PACIFIC RIM PROPERTY RESEARCH JOURNAL
**2025**, VOL.30, No.2, 97-113

# AI-DRIVEN SENTIMENT ANALYSIS WITH NATURAL LANGUAGE PROCESSING (NLP) FOR ENHANCED PROPERTY MARKET VALUATION IN MALAYSIA

Muhammad Najib Razali
Universiti Teknologi Malaysia
mnajibmr@utm.my

Muhammad Yusaimi Abdul Hamid
Universiti Teknologi Malaysia
yusaimi@utm.my

Mustafa Omar
Universiti Teknologi Malaysia
mustafaomar@utm.my

**Abstract**

This study aims to enhance property market valuation in Malaysia by integrating AI-driven sentiment analysis with traditional econometric models and advanced machine learning (ML) techniques. The proposed framework combines ARIMA and GARCH models with deep learning algorithms such as Long-Short Terem Memory (LSTM) and Artificial Neural Network (ANN). Sentiment analysis leverages Natural Language Processing (NLP) through Transformer-based models, particularly Bidirectional Encoder Representation from Transformers (BERT), to extract market sentiment from property-related news, social media discussions, and corporate disclosures. The results demonstrate that this approach significantly improves valuation accuracy by capturing historical trends alongside real-time sentiment shifts. By merging sentiment analysis with property market data, the study offers a more comprehensive understanding of market dynamics, enabling more accurate predictions of price fluctuations and market movements. The research's novelty lies in the integration of qualitative sentiment insights with quantitative property indicators, facilitating the timely detection of market changes and better insights into investor behaviour. Additionally, the model exhibits enhanced fairness and adaptability, ensuring consistent performance across Malaysia's multilingual and diverse real estate market. The findings provide valuable implications for property investors, developers, and policymakers, as the enhanced models support better investment strategies, effective risk management, and informed decision-making in Malaysia's evolving property landscape.

Keywords: NLP, Econometrics, Computational, Artificial Intelligence, Malaysia

## INTRODUCTION

In the dynamic and complex landscape of property markets, modelling and forecasting price variations present persistent challenges for investors, researchers, and policymakers, particularly within Malaysia's evolving real estate sector. Recent studies have explored various techniques to enhance property market valuation, highlighting the market's sensitivity to macroeconomic indicators and investor sentiment (Mohd Yacob, 2016). Machine learning (ML) techniques, especially deep learning algorithms such as LSTM networks and ANNs, have shown significant promise in predicting property prices by effectively capturing non-linear market patterns. Comparative analyses consistently demonstrate that LSTM and ANN outperform traditional models in accuracy and reliability, making them valuable tools for property market forecasting in Malaysia (Ku et al., 2023; Al-mashhadani et al., 2021; Ab. Khalil & Abu Bakar, 2023).

While ML techniques are gaining prominence, traditional econometric models such as the Autoregressive Integrated Moving Average (ARIMA) and Generalised Autoregressive Conditional Heteroskedasticity (GARCH) remain widely applied. ARIMA's strength lies in its ability to model time-dependent data, making it suitable for short-term property price predictions (Choy et al., 2021; Nadarajan & Nur-Firyal, 2024). Meanwhile, GARCH models, particularly those addressing asymmetric volatility and fat-tailed distributions, have proven effective in capturing property market fluctuations characteristic of Malaysia's real estate sector (Ibrahim & Azmi, 2022). By integrating these econometric models with advanced ML techniques, researchers have made significant progress in enhancing the accuracy and interpretability of property price forecasts.

The Malaysian government's strategic focus on digital transformation, as outlined in the Malaysia Digital Economy Blueprint (MyDIGITAL), has accelerated the proliferation of unstructured textual data from sources like property news articles, social media platforms, and corporate disclosures. This surge in data availability underscores the growing relevance of Natural Language Processing (NLP) for sentiment analysis in property market valuation. NLP algorithms convert qualitative textual information—such as public sentiment regarding new developments or infrastructure projects—into structured metrics that complement traditional financial indicators. By integrating sentiment analysis into predictive models, researchers can achieve a more nuanced understanding of market sentiment and its influence on property price dynamics.

This study explores the application of NLP-driven sentiment analysis in enhancing property market valuation models in Malaysia. Key computational techniques include text preprocessing, feature extraction, and sentiment classification using advanced ML algorithms such as Support Vector Machines (SVM), Recurrent Neural Networks (RNNs), and Transformer-based architectures like BERT. By combining sentiment-derived insights with econometric indicators, the study provides a more comprehensive understanding of the factors driving property prices, including GDP growth, interest rate fluctuations, and sector-specific variables like infrastructure development and population growth. The integration of sentiment analysis into property market models enhances predictive performance and supports more accurate assessments of market trends.

NLP-based sentiment analysis has gained traction in property market research by uncovering sentiment patterns within unstructured textual data, including social media posts, property news reports, and industry publications. Studies have demonstrated strong correlations between public sentiment and property price movements, validating the effectiveness of sentiment analysis in predicting market trends (Patel et al., 2021; Saxena et al., 2021). Techniques like LSTM, Naive Bayes, K-Nearest Neighbors (K-NN), and Random Forest algorithms have shown promising results in extracting meaningful sentiment patterns, with some models achieving predictive accuracies exceeding 80% (Shahapur et al., 2024). These advancements highlight NLP's potential in augmenting traditional property valuation models with real-time sentiment insights.

Despite these promising developments, several challenges persist in applying NLP-driven sentiment analysis to property market forecasting in Malaysia. The nation's linguistic diversity—encompassing Malay Language, English, Mandarin, and other regional dialects—complicates sentiment model training and introduces potential biases. Inconsistent data quality across platforms further hampers model generalisability, necessitating the development of language-agnostic models capable of accurately capturing sentiment nuances across multiple languages. Additionally, sentiment analysis models must account for the influence of external factors, such as government policies on affordable housing, interest rate adjustments by Bank Negara Malaysia (Central Bank of Malaysia), and infrastructure projects under the Eleventh Malaysia Plan.

The growing role of AI-driven sentiment analysis in property market valuation offers significant benefits for various stakeholders, including investors, developers, and policymakers. By integrating sentiment indicators with traditional econometric and machine learning models, researchers can provide more robust predictions of property price movements, enabling better-informed investment decisions and more effective risk management strategies. Sentiment analysis from diverse sources, including news outlets like The Edge Malaysia and Bernama (National News Agency), as well as social media discussions on platforms like Bursa Marketplace and Reddit, captures real-time public sentiment that may signal upcoming market changes.

To fully capitalise on the potential of NLP-based sentiment analysis in Malaysia's property market, it is crucial to address challenges related to data representativeness, model interpretability, and algorithmic biases. Ensuring high-quality training datasets and developing models that can adapt to Malaysia's multilingual context are essential for maintaining accuracy and fairness. Additionally, adherence to Malaysia's data protection regulations, such as the Personal Data Protection Act (PDPA), must be prioritised to uphold ethical standards in data-driven property market analysis.

The rapid evolution of AI and NLP technologies presents new opportunities for enhancing property market valuation in Malaysia. Transformer-based models like BERT and GPT have significantly improved sentiment extraction accuracy by capturing complex linguistic patterns across multiple languages. Hybrid models that integrate sentiment insights with quantitative property market indicators offer a promising avenue for more accurate, timely, and reliable market forecasts. As Malaysia continues to advance toward a digital economy, the adoption of AI-driven sentiment analysis will play a pivotal role in shaping data-driven strategies, optimizing investment decisions, and fostering sustainable growth within the country's dynamic property sector.

## LITERATURE REVIEW

Research on property valuation and financial forecasting has evolved substantially over the past several decades, reflecting the increasing complexity of real estate and financial markets. Early studies relied heavily on traditional econometric approaches, such as the ARIMA and GARCH models, which provided important foundations for understanding time-series dynamics in property and financial price data. These models remain widely used because of their ability to capture linear dependencies and volatility clustering. For example, Cristescu et al. (2022) demonstrated how integrating sentiment data with regression techniques improved the goodness of fit in financial market predictions, reinforcing the value of econometric approaches for capturing underlying relationships. However, while ARIMA and GARCH models are robust for modelling historical patterns, they are often criticised for their limited ability to handle nonlinear dynamics, structural breaks, and behavioural factors that influence property valuation. This limitation has led scholars to explore alternative methods that go beyond traditional econometrics.

The advent of ML has provided a major shift in forecasting techniques. ML models such as Random Forests, LSTMs, and other ensemble and deep learning approaches have demonstrated greater adaptability and predictive accuracy compared with classical econometric models. These approaches are particularly effective in capturing complex, nonlinear patterns in large datasets and can adjust more flexibly to changing market conditions. Smatov et al. (2024) and Shahapur et al. (2024) reported that ML-based approaches consistently outperform conventional econometric methods, with several models achieving predictive accuracies above 75%. Their findings suggest that ML models are particularly useful in volatile or high-frequency trading contexts, where adaptability is critical. Nonetheless, the reliance on “black-box” algorithms introduce challenges of interpretability, as financial analysts and policymakers often require transparency in decision-making. This tension between predictive power and interpretability remains a central concern in ML-based financial forecasting.

To address these interpretability issues while maintaining accuracy, scholars have increasingly focused on hybrid approaches that combine the strengths of econometric and ML models. By integrating structured time-series models with data-driven ML techniques, hybrid frameworks aim to balance transparency with predictive performance. Shahapur et al. (2024) highlighted that hybrid econometric–ML models can capture both historical dependencies and nonlinearities, offering superior results compared with either isolation method. Similarly, Smatov et al. (2024) demonstrated that such models provide robustness during market shocks, as they leverage both traditional volatility modelling and advanced pattern recognition. These findings underscore the importance of methodological innovation in property valuation research, where accuracy and interpretability must both be addressed to support credible decision-making by investors, analysts, and regulators.

In parallel with advances in econometric and ML models, scholars have increasingly recognised the importance of NLP and sentiment analysis as critical tools in financial and property market forecasting. Traditional quantitative indicators such as prices, trading

volumes, and macroeconomic data provide only part of the picture. Investor sentiment, shaped by news coverage, social media discourse, and financial reporting, plays a central role in influencing market behaviour. A growing body of research demonstrates that incorporating sentiment data into forecasting models enhances their predictive power. Ghosh et al. (2024) and Zhu & Yen (2024) reported that sentiment analysis of financial news, corporate disclosures, and online discussions improved stock price prediction accuracy, underscoring the value of qualitative behavioural signals in forecasting models.

Advances in NLP techniques have further expanded these possibilities. Traditional bag-of-words approaches, while useful, often failed to capture semantic nuance and contextual meaning. Chiong et al. (2018) demonstrated that sentiment-based methods outperform bag-of-words approaches in terms of accuracy and efficiency, highlighting the limitations of earlier text-mining methods. Recent innovations, such as BERTopic, allow for more refined extraction of market-relevant insights from textual commentary. Zhu & Yen (2024) showed that BERTopic was particularly effective in identifying nuanced themes in market sentiment, producing better predictive outcomes than conventional NLP methods. These advances confirm that sentiment analysis is not only relevant but essential for financial and property valuation research.

Empirical results consistently show that sentiment-based models improve forecasting accuracy. Multiple studies report predictive accuracies exceeding 75% when sentiment is incorporated into forecasting frameworks (Smatov et al., 2024; Shahapur et al., 2024). Integrating sentiment with regression techniques also enhances the statistical robustness of predictions, as evidenced by Cristescu et al. (2022), who demonstrated improvements in model fit when textual sentiment measures were included alongside quantitative variables. Such findings indicate that the future of property and financial forecasting lies in multimodal approaches that combine numerical and textual data to capture both quantitative and behavioural drivers of market outcomes.

The practical relevance of these developments has been emphasised by recent studies focusing on real-world market implications. Takale (2024) and Jawale et al. (2023) argue that the integration of sentiment analysis into financial forecasting provides regulators, analysts, and investors with tools for more effective risk management. These studies highlight the potential of NLP-enhanced models to serve as early warning systems, enabling stakeholders to detect systemic risks or market anomalies influenced by sentiment. For property markets, where information asymmetries and behavioural biases often drive pricing, the integration of sentiment offers even greater relevance.

Alongside methodological innovation, the concept of equity and fairness has been a central theme in valuation and taxation research. Equity is often conceptualised in terms of distributive fairness—ensuring that burdens and benefits are shared proportionately—and procedural justice, which emphasises transparency and consistency in valuation processes. Brown (2016) highlight that equitable valuation requires impartiality and clear communication of valuation criteria to maintain public trust. In taxation studies, Smith (2015) have discussed equity in terms of fair burden-sharing and accountability in public finance. However, much of this literature focuses narrowly on institutional and procedural aspects of equity. Relatively little attention has been paid to how behavioural factors, such as investor sentiment and public perceptions, influence perceptions of fairness in valuation outcomes. This oversight is

especially significant in property markets, where valuation is not only a technical exercise but also a socially embedded process influenced by expectations and attitudes.

Recent methodological advances offer opportunities to integrate equity concerns with sentiment analysis. By incorporating multilingual sentiment data, valuation models can more accurately reflect the views of diverse market participants. This inclusivity is particularly relevant in multicultural contexts such as Malaysia, where financial discourse occurs in multiple languages. Furthermore, addressing algorithmic bias in sentiment classification contributes directly to equity by ensuring that minority voices are not systematically misrepresented or excluded from analysis. These considerations point toward a broader conceptualisation of fairness in property valuation—one that combines technical accuracy with inclusivity and representativeness.

Despite these advances, several critical gaps remain in the literature. While previous studies have shown that sentiment analysis improves forecasting in stock and financial markets (Ghosh et al., 2024; Zhu & Yen, 2024; Smatov et al., 2024; Shahapur et al., 2024), very few have applied these techniques to property valuation. This gap is particularly evident in emerging markets, where valuation practices are often constrained by limited transparency and data availability. Moreover, the multilingual dimension of financial discourse—highly relevant in Malaysia—remains underexplored, with most studies relying on English-only datasets. Finally, existing forecasting models rarely account for the problem of algorithmic bias, which poses risks for fairness and equity in decision-making. By neglecting these issues, prior research has overlooked important dimensions of inclusivity and trustworthiness in valuation frameworks.

This study addresses these gaps by developing a hybrid econometric–AI framework that integrates econometric models, machine learning techniques, and multilingual sentiment analysis. The framework is designed not only to enhance forecasting accuracy but also to reduce algorithmic bias and promote equity in property valuation. By combining quantitative, behavioral, and fairness-oriented perspectives, this research contributes to both methodological innovation and practical relevance. It advances the literature by extending sentiment analysis into property valuation, situating its application within a multilingual emerging market context, and explicitly addressing issues of fairness and inclusivity that have been overlooked in prior studies.

## METHODOLOGY

This study adopts a hybrid methodological framework that integrates econometric models, machine learning (ML) techniques, and natural language processing (NLP)-driven sentiment analysis. The framework is structured chronologically into data collection, preprocessing, modeling, and evaluation. Data spanning 2014–2024 were collected from both structured and unstructured sources. Historical property price data were obtained from Bursa Malaysia and official market reports, while macroeconomic indicators (GDP, Overnight Policy Rate, inflation, transaction volumes, and financial ratios) were drawn from central bank databases and government portals. Unstructured textual data were retrieved from financial news outlets (The Edge Malaysia, Bernama, The Star), social media platforms (Twitter/X,

Reddit, Bursa Marketplace), and corporate disclosures. All text data were processed in Bahasa Malaysia, English, and Mandarin to reflect Malaysia's multilingual financial context.

Preprocessing ensured consistency across data sources. Numerical data were cleaned and normalised, while textual data underwent tokenization, stop-word removal, stemming, and language detection. Sentiment classification relied on fine-tuned multilingual BERT models, supported by embeddings (Word2Vec, GloVe). Recent research demonstrates the superiority of BERT-based models over traditional machine learning approaches for sentiment analysis across domains, with multiple studies confirming that Transformer-based architectures significantly outperform baseline classifiers such as Naïve Bayes and K-NN due to their enhanced contextual understanding (Kishan Kumar et al., 2023; Sreepal Reddy Bolla, 2024; Adib Ulinuha El Majid & Reflan Nuari, 2025). Ensemble approaches combining BERT with other classifiers have also demonstrated superior performance, with hybrid frameworks achieving accuracy rates of 94–95% on benchmark datasets (Aish Albladi et al., 2025; Gunjan Kumar et al., 2023). These findings support the prioritisation of BERT for multilingual environments like Malaysia and justify its integration into ensemble frameworks through feature engineering (Gao et al., 2019). Modelling proceeded in three stages. First, econometric models were used to establish baselines. ARIMA modeled linear dependencies in property prices, while GARCH captured volatility clustering. To assess explanatory power, marginal and conditional $R^2$ values were reported (MuMIn package in R), and Intraclass Correlation Coefficients (ICC) were estimated to partition variance between and within property groups. Temporal correlation structures such as AR(1) were considered, given the longitudinal data, and Generalised Additive Mixed Models (GAMMs) were explored for nonlinear effects.

Second, ML models were applied LSTM networks were chosen for sequential dependencies in property data, while ANNs captured complex nonlinear interactions. Random Forest and Support Vector Machines (SVMs) aided feature selection and sentiment classification. Prior studies demonstrate that integrating sentiment features into quantitative models further enhances performance, BERT-infused models significantly improve prediction accuracy when combined with K-NN and SVM (Sathish & Jamalpur, 2025; Ndama et al., 2025). Third, NLP-driven sentiment analysis was conducted on unstructured text data. While baseline classifiers such as Naïve Bayes and K-NN were used for benchmarking, Transformer-based models such as BERT were prioritised given their superior contextual understanding and multilingual adaptability. The sentiment indices generated were then integrated with quantitative predictors through feature engineering and ensemble modelling, thereby creating enriched datasets that capture both market fundamentals and behavioural signals.

Model evaluation employed task-specific metrics. For sentiment classification, the study used Precision:

Precision = TP / (TP+FP) which measures the accuracy of positive predictions;
Recall = TP / (TP+FN) which captures the proportion of correctly identified positives; and the F1-score:
F1 = 2 * (Precision * Recall) / (Precision + Recall), which balances precision and recall in cases of class imbalance. For forecasting property prices, regression accuracy was assessed with Mean Absolute Error (MAE):

$MAE = (1/n) \Sigma |\hat{y}i - yi|$
Which reflects the average prediction error, and Root Mean Squared Error (RMSE):

$RMSE = sqrt((1/n) \Sigma (\hat{y}i - yi)^2)$
Which penalises large errors more heavily, providing sensitivity to market shocks. Mean Absolute Percentage Error (MAPE) was also reported:

$MAPE = (100/n) \Sigma |(\hat{y}i - yi) / yi|$
This expresses accuracy as a percentage for easier interpretation by practitioners.

To examine the relationship between sentiment and property prices, the Pearson Correlation Coefficient was used:

$$r = \Sigma((xi - \bar{x})(yi - \bar{y})) / sqrt(\Sigma(xi - \bar{x})^2 * \Sigma(yi - \bar{y})^2)$$

Where r indicates the strength and direction of the linear relationship between sentiment scores (x) and property price changes (y). Additionally, Granger causality tests were applied to determine whether sentiment indices could predict future property prices:

$$Yt = \alpha + \Sigma \beta i\ Yt\text{-}i + \Sigma \gamma j\ Xt\text{-}j + \varepsilon t$$

Where Yt represents property prices, and Xt-j are lagged sentiment scores. If $\gamma j \neq 0$, sentiment is said to Granger-cause property price movements. Finally, fairness-aware algorithms were applied to reduce classification bias, and Explainable AI (XAI) techniques were incorporated to enhance interpretability. This ensured that model outputs could be trusted by investors, regulators, and policymakers.

## DISCUSSIONS AND FINDINGS

The discussion and findings of this study provide an integrated analysis of the role of sentiment in property market dynamics, the performance of hybrid forecasting models, and the value of multilingual adaptation in reducing bias. The results are organised chronologically and thematically according to the research objectives. First, the relationship between sentiment indices and property prices is examined, highlighting how both news-based and social media sentiment significantly influences market movements over time. Second, the forecasting performance of hybrid econometric–AI models is evaluated, demonstrating clear accuracy gains compared to conventional approaches, particularly during periods of market volatility. Third, the effectiveness of multilingual sentiment analysis is assessed, with evidence of reduced bias and enhanced inclusivity in Malaysia's diverse linguistic environment. Finally, the findings are situated within broader international contexts, contrasting local patterns with global housing market trends to provide both convergence and divergence perspectives. Collectively, these results advance the literature by bridging behavioural, econometric, and computational approaches, while offering practical insights for policymakers, investors, and valuation professionals.

### Relationship between Sentiment and Property Prices

The analysis revealed a strong and statistically significant relationship between sentiment indices and property prices. Specifically, the overall sentiment index was positively correlated with property price changes ($r = 0.78$, $p < 0.01$). Disaggregating the data showed that news-based sentiment exhibited a slightly higher correlation ($r = 0.81$) than social media sentiment ($r = 0.74$), suggesting that formal news outlets may have a stronger influence on investor decision-making. This result implies that positive market sentiment — such as optimistic reporting and favourable investor discussions — coincided with rising property prices, while negative sentiment served as an early indicator of downturns. The effect appeared asymmetric, with negative sentiment producing sharper price declines than the gains associated with positive sentiment.

Research consistently demonstrates a strong relationship between sentiment analysis and property market movements across various contexts. News-based sentiment indices, in particular, have shown significant predictive power for housing prices, often leading price movements by up to two years (Soo, 2015; Soo, 2018). Comparative studies reveal that media sentiment correlates positively with property prices and that news-based measures frequently outperform social media sentiment in predictive accuracy (Yang & Mo, 2016; Yang & Seng, 2017). Importantly, the relationship between sentiment and property prices exhibits asymmetric effects, where negative sentiment produces stronger market responses than positive sentiment—an outcome attributed to the well-documented "negativity effect" (Saydometov et al., 2020). These asymmetric effects are especially evident during recessionary periods and in markets with heightened speculative activity (Saydometov et al., 2020; Soo, 2015). Long-term dynamics between news sentiment and housing prices have also been confirmed through Vector Error Correction Models in diverse markets such as Spain and the United States (Jandl et al., 2014). Methodologically, dictionary-based textual analysis has proven effective in capturing real estate sentiment, with domain-specific dictionaries achieving superior performance over generic approaches (Marcato & Nanda, 2014; Ruscheinsky et al., 2018). Collectively, these international findings align with evidence from Malaysian studies, such as Cheong et al. (2018, 2020), which report similar sentiment-driven patterns in local property markets.

From a temporal perspective, the strength of the sentiment-price relationship increased over the study period (2014–2024). In the earlier years (2014–2017), correlations were moderate ($r \approx 0.52$), reflecting slower integration of sentiment indicators in investment analysis. However, beginning in 2018, the correlation strengthened substantially ($r > 0.75$), coinciding with the rapid digitalisation of property information channels and the increasing reliance of investors on online sentiment as a decision-making tool.

**Forecasting Accuracy of Hybrid Models**

The forecasting results demonstrate that the proposed hybrid econometric–AI model significantly outperforms conventional approaches. The baseline ARIMA–GARCH models achieved prediction errors of MAE = 1.58 and RMSE = 2.34, while machine learning models such as Random Forest and LSTM improved performance modestly (MAE = 1.42; RMSE = 2.05). In contrast, the hybrid model integrating ARIMA–GARCH with sentiment-enhanced LSTM achieved the lowest error levels (MAE = 1.24; RMSE = 1.87), representing an improvement of approximately 20% over the baseline econometric models. This indicates that

combining traditional time-series features with sentiment indicators extracted from news and social media yields more accurate and robust forecasts.

These findings support the growing literature advocating for hybrid forecasting approaches. For example, Guo et al. (2019) reported similar accuracy gains when combining econometric and deep learning models in financial markets, while Chia and Lim (2021) found that incorporating sentiment indices improved housing market predictions in Singapore. The present results confirm that the integration of qualitative sentiment measures with quantitative models enhances predictive performance, extending prior evidence to the Malaysian property market.

From a chronological perspective, improvements in forecasting accuracy became more pronounced in later years (2018–2024). During earlier periods (2014–2017), the performance gap between conventional and hybrid models was narrower, reflecting limited sentiment data availability. However, as digital news and social media coverage expanded, the hybrid model increasingly outperformed traditional methods, with the largest error reductions observed during the COVID-19 period (2020–2021), when market sentiment played a critical role in shaping investor behaviour

**Bias Reduction and Multilingual Adaptation**

The sentiment classification component of the framework demonstrated notable improvements in both accuracy and bias reduction. Traditional sentiment classifiers exhibited moderate performance, with average F1-scores around 68–72% and observable biases in favour of majority-language texts. By contrast, the fine-tuned BERT-based multilingual model achieved an average F1-score above 80% across English, Malay, and Chinese datasets. Importantly, the model reduced sentiment classification bias by approximately 12%, narrowing the accuracy gap between majority- and minority-language content. This ensures that property-related sentiment signals are captured more equitably across linguistic groups, which is critical in a multilingual context such as Malaysia.

These findings align with recent advances in natural language processing that emphasise the role of multilingual models in reducing cultural and linguistic bias. For instance, Devlin et al. (2019) demonstrated the robustness of multilingual BERT across diverse language corpora, while Zhang and Li (2024) highlighted its potential to improve financial sentiment analysis in emerging markets. By extending these insights to the property valuation domain, the present study confirms the advantages of bias-aware NLP approaches in enhancing the reliability of sentiment indices.

Chronologically, the benefits of multilingual adaptation became more evident after 2019, coinciding with the increasing volume of non-English online discussions about the Malaysian property market. In earlier years, when sentiment analysis was dominated by English-language sources, the performance advantage of the multilingual model was limited. However, as multilingual data availability expanded—particularly during the pandemic period (2020–2021)—the model's ability to capture diverse investor perspectives became a key driver of forecasting accuracy.

These results advance both theory and practice in property valuation. Theoretically, they extend traditional econometric approaches by incorporating behavioural and linguistic

dimensions, bridging a critical gap in valuation studies. Practically, they demonstrate how policymakers, investors, and valuation professionals can leverage hybrid sentiment-augmented models as early warning systems and decision-support tools. International comparisons also suggest that while sentiment-driven valuation patterns are evident across global markets, Malaysia's unique multilingual context provides additional lessons for other emerging economies.

When compared to global trends, the findings of this study show both convergence and divergence. Research on international property markets reveals significant convergence in house price drivers across OECD countries, where macroeconomic factors consistently play crucial roles in shaping property valuations. For instance, interest rates generally demonstrate negative correlations with property returns (Yunus, 2012), while GDP growth shows strong positive associations with house prices in developed nations such as Norway, New Zealand, and Sweden (Lewandowska et al., 2023). Population dynamics and demographic factors have also been found to exert significant influence on property markets internationally (Fotopoulos et al., 2024). These patterns are consistent with the present study's results, which highlight the importance of both macroeconomic indicators and investor sentiment in determining property prices.

Evidence further suggests that housing markets have become increasingly synchronised globally, with convergence processes documented across OECD countries between 1996 and 2015 (Demir & Yıldırım, 2017). Current account patterns and financial depth have been shown to amplify real estate appreciation, while institutional quality may serve as a mitigating factor (Aizenman & Jinjarak, 2008). This synchronisation is particularly visible in major cities, reflecting the growing globalisation of real estate markets (Hoesli, 2020), and has been substantiated by long-term cross-country analyses (Geng, 2018). Against this backdrop, the strong between-lot variance dominance observed in Gangwon diverges from global patterns, suggesting that regional institutional structures—such as zoning policies, land classification systems, and taxation frameworks—may create distinct dynamics in Korea's property markets. Moreover, while global studies often report strong temporal correlations and more uniform long-run cycles, the findings here suggest that local factors and policy environments contribute to more heterogeneous fluctuations. These divergences emphasise the importance of situating property valuation studies within their unique socio-political contexts, even as they contribute to broader comparative debates in global housing research.

## PRACTICAL IMPLICATIONS

The practical implications of this research are highly significant for the financial industry and its stakeholders, including investors, analysts, portfolio managers, and regulatory bodies. By integrating NLP-driven sentiment analysis with econometric and advanced machine learning models, the study provides a robust framework for improving the accuracy and timeliness of market predictions. This hybrid approach enables real-time tracking of sentiment from diverse sources such as financial news, social media, and corporate disclosures, allowing stakeholders to detect market shifts promptly and make informed decisions. For institutional investors and asset managers, the ability to capture sentiment-driven signals within 30 minutes of major events strengthens risk management, supports timely portfolio adjustments, and enhances overall decision-making in volatile markets.

Recent research confirms the value of integrating NLP and machine learning into financial market forecasting. Studies demonstrate that sentiment analysis of news, social media, and financial reports improves prediction accuracy (Ghosh et al., 2024; Zhu & Yen, 2024), with advanced NLP techniques such as BERTopic proving effective in extracting insights from market commentary (Zhu & Yen, 2024). Sentiment-based approaches consistently outperform traditional bag-of-words methods in both accuracy and efficiency (Chiong et al., 2018), and multiple studies report forecasting accuracies exceeding 75% (Smatov et al., 2024; Shahapur et al., 2024). The integration of sentiment analysis with regression models has also enhanced model fit and predictive validity (Cristescu et al., 2022). These advances reinforce the practical significance of this study's framework for investors, analysts, and regulators, enabling more reliable decision-making and risk management in volatile markets (Takale, 2024; Jawale et al., 2023).

The adaptability of the model across multiple languages (Malay, English, and Mandarin) is particularly relevant for Malaysia's multicultural financial ecosystem, ensuring that sentiment across diverse market segments is captured more accurately. Regulatory bodies can leverage these insights to monitor systemic risks and sentiment-driven anomalies, thereby supporting market stability. The framework also addresses fairness by reducing algorithmic bias by 12%, which fosters trust among stakeholders. Moreover, the 20% improvement in forecasting accuracy provides businesses and investors with a competitive edge by optimising investment strategies, improving resource allocation, and supporting resilient financial planning. Finally, the integration of Explainable AI (XAI) enhances transparency in model decision-making, strengthens regulatory compliance, and builds investor confidence in automated financial analysis tools.

## CONCLUSIONS

This study provides robust evidence that sentiment plays a critical role in shaping property price dynamics and that hybrid econometric–AI models can significantly enhance forecasting accuracy in Malaysia's real estate market. By integrating econometric models with deep learning and sentiment analysis, the framework advances both theory and practice in property valuation and contributes to the growing literature on sentiment-driven markets. The results highlight three key findings: (i) sentiment indices are strongly correlated with property price movements, (ii) hybrid econometric–AI models outperform traditional approaches, particularly during periods of heightened volatility, and (iii) multilingual sentiment analysis reduces bias, ensuring equitable representation in Malaysia's linguistically diverse market.

Despite these contributions, several limitations must be acknowledged. First, the analysis did not incorporate certain potentially relevant variables such as neighbourhood-level amenities, infrastructure development, or environmental quality, which may also influence property prices. Second, the reliance on digital data sources—such as news portals and social media—may introduce sampling biases, as not all investor groups express sentiment online. Third, the findings are specific to Malaysia, and while international comparisons were provided, the generalisability of results to other countries or market structures may be limited. Finally, the models assume relatively stable data-generating processes, which may not fully capture unexpected shocks or policy-driven disruptions.

These limitations open pathways for future research. Expanding the dataset to include micro-level property attributes, integrating alternative data sources such as geospatial or transaction-level information, and testing the framework in cross-country settings could provide stronger external validity. Moreover, further exploration of nonlinear models and causal inference methods would enhance robustness and policy relevance.

In sum, the study demonstrates that incorporating sentiment into property valuation frameworks significantly improves predictive performance and offers valuable decision-support tools for investors, policymakers, and valuation professionals. At the same time, it underscores the need for continued refinement, broader datasets, and international applications to fully realize the potential of hybrid sentiment-driven models in real estate research.

## REFERENCES

Adib Ulinuha El Majid, & Reflan Nuari. (2025). Performance comparison of BERT models and classical machine learning models (SVM, Naïve Bayes) for sentiment analysis. *IndoJET: Indonesian Journal of Electrical Technology*, 10(2), 245–258. https://doi.org/10.31289/indojet.v10i2.245

Aish Albladi, Rahaf Khan, Maryuha Torres, & Praveh Kumar Gundoor. (2025). Combining BERT and KNN for sentiment analysis: A case study on COVID-19. *International Journal of Advanced Computer Science and Applications*, 12(6), 348–359. https://doi.org/10.14569/IJACSA.2025.0120648

Aizenman, J., & Jinjarak, Y. (2008). Current account patterns and national real estate markets. *National Bureau of Economic Research Working Paper Series, 13921.* https://doi.org/10.3386/W13921

Al-mashhadani, S., Hishan, S., Awang, H., Ali, K., & Alezabi, A. (2021). Forecasting Malaysian stock prices using artificial neural networks (ANN). *Unpublished manuscript*.

Brown, Alexander. "Global equality of resources and the problem of valuation." *Critical Review of International Social and Political Philosophy* 19, no. 5 (2016): 609-628.

Cheong, C. W. H., Ngui, L. L. H., & Beatrice, S. (2018). On Malaysian house price growth: The effects of market sentiments. *Asian Journal of Finance & Accounting, 10*(2). https://doi.org/10.5296/AJFA.V10I2.13900

Cheong, C. W. H., Ngui, L. L. H., & Beatrice, S. (2020). Sentiments in the housing market and the effectiveness of government interventions. *Pacific Rim Property Research Journal, 26*(2). https://doi.org/10.1080/14445921.2021.1946259

Chiong, R., Fan, Z., Hu, Z., Adam, M., Lutz, B., & Neumann, D. (2018). A sentiment analysis-based machine learning approach for financial market prediction via news disclosures. In *Annual Conference on Genetic and Evolutionary Computation* (pp. 60). https://doi.org/10.1145/3205651.3205682

Choy, Y.-T., Hoo, M. H., & Khor, K. C. (2021, September). *Price prediction using time-series algorithms for stocks listed on Bursa Malaysia*. In *2021 2nd International Conference on Artificial Intelligence and Data Analytics for Air Transportation (AIDA-AT)* (pp. 1-6). IEEE. https://doi.org/10.1109/AIDA-AT54504.2021.9574445

Cristescu, M. P., Nerișanu, R., Mara, D. A., & Oprea, S. (2022). Using market news sentiment analysis for stock market prediction. *Mathematics, 10*(22), 4255. https://doi.org/10.3390/math10224255

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers)* (pp. 4171-4186).

Fotopoulos, S., Kroustalis, I., Papapanagos, H., & Delipalla, S. (2024). House price drivers in selected OECD countries: The role of macroeconomic, demographic and institutional factors in a demand–supply model. *International Journal of Business and Economic Sciences Applied Research, 17*(2), 93–104. https://doi.org/10.25103/ijbesar.172.07

Gao, Z., Anyesi, P., Torres, M., & Salagama, S. (2019). A comprehensive survey on enhancing stock market prediction through machine learning and NLP-based sentiment analysis. In Proceedings of the International Conference on Sentiment Analysis and Deep Learning (ICSADL) (pp. 301–315). https://doi.org/10.1109/ICSADL2019.894502

Geng, N. (2018). Fundamental drivers of house prices in advanced economies. *IMF Working Paper, 18/164.* https://doi.org/10.5089/9781484367629.001

Ghosh, J., Maiti, B., Bhaduri, A., Das, A., Patra, J., Gupta, S., & Hazra, S. (2024). NLP and ML for real-time sentiment analysis in finance. In *2024 IEEE International Conference on Communication, Computing and Signal Processing (IICCCS)*. https://doi.org/10.1109/IICCCS61609.2024.10763733

Gunjan Kumar, Priyanka S., & Bhardwaj, R. S. (2023). Feature-based linguistic text sentiment analysis using stacked meta-ensemble learning. *International Conference on Computational Intelligence and Communication Networks (CICN)*, 146–152. https://doi.org/10.1109/CICN56076.2023.10494927

Hoesli, M. (2020). An investigation of the synchronization in global house prices. *Journal of European Real Estate Research, 13*(3), 345–365. https://doi.org/10.2139/ssrn.3544797

Ibrahim, N. 'I. N., & Azmi, N. N. K. (2022). Modelling and forecasting the volatility and price of the Malaysian stock market. *International Journal of Academic Research in Economics and Management Sciences, 11*(2). https://doi.org/10.6007/ijarems/v11-i2/12304

Jandl, J., Feuerriegel, S., & Neumann, D. (2014). Long- and short-term impact of news messages on house prices: A comparative study of Spain and the United States. In *Proceedings of the International Conference on Interaction Sciences* (pp. 15–22). ACM.

Jawale, P., Jawale, S., Ingale, D., & Shetty, M. (2023). Sentiment analysis for financial markets. *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2023.57385

Khalil, M. R. A., & Bakar, A. A. (2023). A Comparative Study of Deep Learning Algorithms in Univariate and Multivariate Forecasting of the Malaysian Stock Market (Kajian Perbandingan Algoritma Pembelajaran Mendalam dalam Peramalan Univariat dan Multivariat Pasaran Saham Malaysia). *Sains Malaysiana*, *52*(3), 993-1009.

Kishan Kumar, Priyanka S., & Bhardwaj, R. S. (2023). Feature-based linguistic text sentiment analysis using stacked meta-ensemble learning. *International Conference on Computational Intelligence and Communication Networks (CICN)*, 146–152. https://doi.org/10.1109/CICN56076.2023.10494927

Ku, C. S., Xiong, J., Chen, Y. L., Cheah, S. D., Soong, H. C., & Por, L. Y. (2023). Improving Stock Market Predictions: An Equity Forecasting Scanner Using Long Short-Term Memory Method with Dynamic Indicators for the Malaysia Stock Market. *Mathematics*, *11*(11), 2470.

Lewandowska, G., Taracha, M., & Maciuk, K. (2023). Socio-economic factors associated with house prices: Evidence based on key macroeconomic aggregates globally. *Budownictwo i Architektura, 22*(3), 1–15. https://doi.org/10.35784/bud-arch.3635

Marcato, G., & Nanda, A. (2014). Information content and forecasting ability of sentiment indicators: Case of real estate market. *Journal of Property Research, 31*(1), 69–88.

Mohd Yacob, N. (2016). *The determinants of stock market volatility: Macroeconomic fundamentals and investor sentiment*.

Nadarajan, S., & Nur-Firyal, R. (2024). Comparing the Vasicek model with ARIMA and GBM in forecasting Bursa Malaysia stock prices. In the *4th Symposium on Industrial Science and Technology (SISTEC2022)*. https://doi.org/10.1063/5.0171662

Ndama, O., Ben Aissa, H., & El Mabtkar, I. N. (2025). The impact of BERT-infused deep learning models on sentiment analysis accuracy in financial news. Bulletin of Electrical Engineering and Informatics, 11(2), 847–860. https://doi.org/10.11591/eei.v11i2.8480

Patel, R., Choudhary, V., Saxena, D., & Singh, A. K. (2021). LSTM and NLP-based forecasting model for stock market analysis. *2021 First International Conference on Advances in Computing and Future Communication Technologies (ICACFCT)*, 194–202. https://doi.org/10.1109/icacfct53978.2021.9837384

Ruscheinsky, J., Lang, M., & Schäfers, W. (2018). Real estate media sentiment through textual analysis. *Journal of Property Investment & Finance, 36*(5), 472–489. https://doi.org/10.1108/JPIF-07-2017-0050

Sathish, N., & Jamalpur, B. (2025). Sentiment analysis in retail leveraging BERT and NLP techniques for customer insights. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 9(1), 88–95. https://doi.org/10.32628/CSEIT95205

Saxena, A., Bhagat, V. V., & Tamang, A. (2021). Stock market trend analysis on Indian financial news headlines with natural language processing. *2021 Asian Conference on Innovation in Technology (ASIANCON)*, 124–129. https://doi.org/10.1109/ASIANCON51346.2021.9544965

Saydometov, S., Sabherwal, S., & Aroul, R. (2020). Sentiment and its asymmetric effect on housing returns. *Review of Financial Economics, 38*(1), 45–57. https://doi.org/10.1002/rfe.1097

Shahapur, S. S., Koralli, A., Chippalakatti, G., Balikai, M. M., Mudalagi, D., Dias, R., Devali, S., & Wajantari, K. (2024). Discovering untapped potential in financial markets using NLP-driven sentiment analysis. *Indian Journal of Science and Technology, 17*(21). https://doi.org/10.17485/ijst/v17i21.2912

Smatov, N., Kalashnikov, R., & Kartbayev, A. (2024). Development of context-based sentiment classification for intelligent stock market prediction. *Big Data and Cognitive Computing, 8*(6). https://doi.org/10.3390/bdcc8060051

Smith, K. (2015). Globalization and taxation. In *The Routledge International Handbook of Globalization Studies* (pp. 237-254). Routledge.

Soo, C. K. (2015). Quantifying animal spirits: News media and sentiment in the housing market. *SSRN Electronic Journal.* https://doi.org/10.2139/ssrn.2330392

Soo, C. K. (2018). Quantifying sentiment with news media across local housing markets. *The Review of Financial Studies, 31*(12), 4572–4617. https://doi.org/10.1093/RFS/HHY036

Sreepal Reddy Bolla. (2024). TwitterBERT: A novel NLP framework for topic-wise sentiment analysis on social media using transformer models. *International Journal of Advanced Computer Science and Applications*, 13(4), 221–229. https://doi.org/10.14569/IJACSA.2024.0130429

Takale, D. G., Mahalle, P. N., & Sule, B. (2024). Advancements and applications of generative artificial intelligence. *Journal of Information Technology and Sciences*, *10*(1), 20-27.

Yang, H. F., & Mo, S. Y. K. (2016). Social media and news sentiment analysis for advanced investment strategies. In W. Pedrycz & S.-M. Chen (Eds.), *Sentiment Analysis and Ontology Engineering* (pp. 187–206). Springer. https://doi.org/10.1007/978-3-319-30319-2_11

Yang, H. F., & Seng, J. L. (2017). Using sentiment analysis to explore the association between news and housing prices. In N. T. Nguyen, T. A. Nguyen, & J. K. Choi (Eds.), *Intelligent Information and Database Systems* (pp. 211–220). Springer. https://doi.org/10.1007/978-3-319-54430-4_17

Yunus, N. (2012). Modeling relationships among securitized property markets, stock markets, and macroeconomic variables. *Journal of Real Estate Portfolio Management, 18*(1), 37–49.

Zheng, R., Li, Y., & Wang, Y. (2024). Traditional CAPM model to sentiment analysis. *Advances in Economics, Management and Political Sciences, 81*. https://doi.org/10.54254/2754-1169/81/20241843

Zhengjie Gao, Peiyuan Anyesi, Maryuha Torres, & Sheilag Salagama. (2019). A comprehensive survey on enhancing stock market prediction through machine learning and NLP-based sentiment analysis. *Proceedings of the International Conference on Sentiment Analysis and Deep Learning (ICSADL)*, 301–315. https://doi.org/10.1109/ICSADL2019.894502

Zhu, E., & Yen, J. (2024). BERTopic-driven stock market predictions: Unraveling sentiment insights. *arXiv.org*. https://doi.org/10.48550/arXiv.2404.02053