FORECASTING NEW ZEALAND HOUSING DEMAND: COMPARING ELASTIC NET, XGBOOST AND RNN MODELS

BINH THANH DANG [1*], CHUONG BUI VIET [2]

[1] *Faculty of Information Technology, Industrial University of HCM City, Ho Chi Minh City, Vietnam,*

[2] *School of Information Technology, Whitecliffe College, Wellington, New Zealand*

** Corresponding author: binhdang@iuh.edu.vn*

DOIs: https://www.doi.org/10.46242/jstiuh.v77i5.5609

Abstract. Forecasting housing demand has been a prevalent research focus globally, primarily employing traditional econometric methods. However, the application of machine learning in this domain remains limited, particularly in the New Zealand context. This study addresses this gap by implementing Elastic Net, XGBoost, and Recurrent Neural Network models to predict residential housing demand in New Zealand using historical economic and demographic data from 1995. The models were evaluated using a comprehensive framework of six complementary metrics (R^2, SMAPE, MAE, RMSE, MBE, and MASE), with the RNN model achieving the highest accuracy. Results demonstrate that machine learning algorithms significantly enhance housing demand forecasting, with temporal models outperforming traditional approaches. The analysis of feature importance identified CPI, construction investment, import values, and unemployment as key drivers, while demographic factors showed limited impact on housing demand. These findings provide valuable insights for policymakers and construction firms addressing New Zealand's housing challenges. Future research should expand dataset dimensions and improve model interpretability.

Keywords. Machine Learning, Neural Networks, Housing Demand, Forecasting Models, Time Series Analysis

1 INTRODUCTION

New Zealand faces a critical housing crisis, where construction—the nation’s fifth-largest industry, contributing 6.3% to GDP [1]—struggles to meet growing demand, resulting in an unaffordable market [2]. The housing market faces two predominant concerns: rising prices and the imbalance between demand and supply [3]. While housing price prediction has been extensively researched, studies focusing on residential construction demand remain scarce, particularly in New Zealand, exacerbating the challenges of addressing supply shortages that adversely affect vulnerable populations mentally and physically [4].

Accurate forecasting of housing demand is essential to guide policymakers, suppliers, developers, and contractors in planning and mitigating these issues. However, traditional econometric methods, widely used for housing demand forecasting globally [5], often fail in dynamic markets like New Zealand. These methods struggle to capture complex, non-linear relationships between economic and demographic factors and require significant human judgment to model temporal dynamics, limiting their effectiveness [6].

This study proposes using machine learning (ML) to forecast housing demand in New Zealand, leveraging ML’s ability to automatically identify non-linear correlations and temporal patterns without human intervention [6]. This study applies advanced ML algorithms (Elastic Net, XGBoost, RNN) to data from the Reserve Bank of New Zealand to identify key demographic and economic indicators and compare model accuracy, aiming to determine which model delivers the best overall performance, with high R^2 and low error values across the selected metrics.

The rest of this paper is structured as follows. Section 2 provides a review of existing literature on machine learning techniques and housing demand prediction. Section 3 outlines the methodology used in the study. Section 4 discusses the results and offers a comparative analysis of model performance. Finally, the paper concludes with key findings and directions for future research.

2 LITERATURE REVIEW

2.1 Machine Learning Foundations

Machine learning represents an application of computer science that employs algorithms to develop statistical analysis models capable of automatic processing without continuous human intervention. These models generate trending patterns from input data that can predict future events or unknown indices [7].

ML algorithms fall into three main categories: supervised, unsupervised, and reinforcement learning. This research focuses specifically on supervised ML [8], which utilizes labeled datasets with predetermined features. In supervised learning, the researcher selects input and output features and determines data allocation for training and testing purposes. The models then identify patterns within the dataset and generate predictions for unknown values. Supervised ML encompasses two primary categories: regression and classification.

2.1.1 Regression Algorithms

Regression models [9] are trained on numerical datasets to uncover patterns in input data and predict continuous dependent variable values. This study employs the following regression algorithms:

2.1.1.1 Elastic Net

Elastic Net [10] integrates the advantages of both Lasso and Ridge Regression, with a penalty function that combines both approaches:

$$p(\beta) = \lambda_1 \textstyle\sum_{j=1}^{m} \beta_j^2 + \lambda_2 \sum_{j=1}^{m} |\beta_j| \tag{1}$$

This algorithm evaluates parameters of correlated variables collectively, determining whether to retain or remove them based on their collective impact on the target variable, thereby enhancing prediction accuracy.

2.1.1.2 XGBoost

XGBoost (eXtreme Gradient Boosting) [11] employs decision tree methods and gradient boosting frameworks [12]. It operates by minimizing residual sums with hyperparameters to prevent overfitting:

$$Obj = \sum_{i=1}^{n} (y_i - \hat{y_i})^2 + \gamma T + \frac{1}{2}\lambda\omega^2 \tag{2}$$

where γ represents the regularization parameter for leaf numbers, T denotes tree leaf count, λ is the leaf weight parameter, and ω is individual leaf weight.

XGBoost begins with an initial random prediction and sequentially builds trees, with each subsequent tree aiming to minimize preceding errors. The final output combines the initial prediction with outputs from all trees, scaled by a learning rate:

$$\hat{y_i} = \widehat{y_i^0} + \theta \sum_{m=1}^{M} f_m(x_i) \tag{3}$$

where θ represents the learning rate controlling new tree contributions to the final prediction.

2.1.1.3 Recurrent Neural Networks (RNN)

RNNs [13], a fundamental artificial neural network (ANN) model, incorporate feedback loops where previous hidden neuron outcomes inform subsequent neurons, aiming to minimize output-target differences. The hidden state $\mathrm{h_t}$ is expressed as:

$$\mathrm{h_t} = \tanh(\mathrm{W_h x_t} + \mathrm{U_h h_{t-1}} + \mathrm{b_h}) \tag{4}$$

where $\mathrm{W_h}$ and $\mathrm{U_h}$ are weight matrices, $\mathrm{x_t}$ is the input at time step t, $\mathrm{h_{t-1}}$ is the hidden state of the previous time step, $\mathrm{b_h}$ is the bias parameter, and $\tanh$ serves as the activation function.

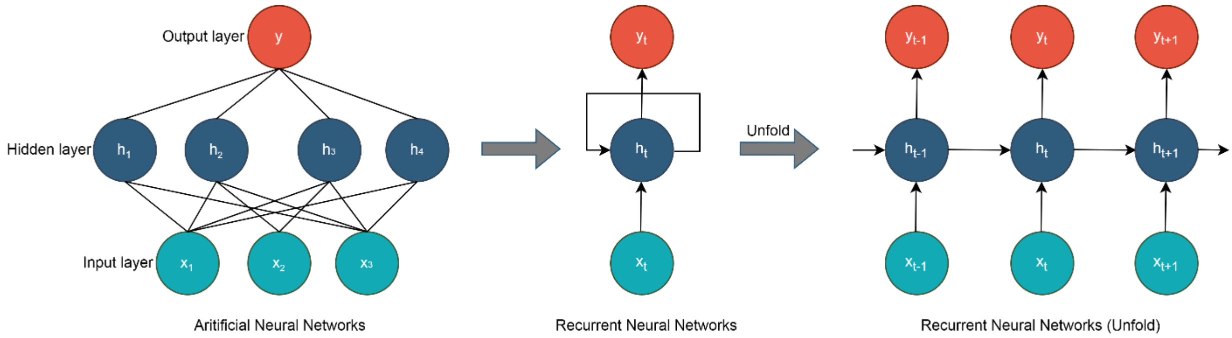

Figure 1: Structure of artificial neural networks and recurrent neural networks

2.2 Related Research

2.2.1 Traditional Econometric Methods

Before the widespread adoption of ML predictive models, researchers employed various traditional techniques such as AutoRegressive Integrated Moving Average (ARIMA), Box-Jenkins, and panel data analysis. These methods continue to be used alongside ML models. Studies by [14] and [15] developed models forecasting private housing demand in various Turkish cities using econometric methods. Both studies were limited by the number of independent variables they could incorporate. Additionally, significant human intervention in selecting statistical functions limited their ability to handle complex models, a disadvantage that machine learning approaches typically overcome.

Goh [6] compared the accuracy of ANN and Box-Jenkins forecasting models for housing demand in Singapore from 1975 to 1994. The ML model demonstrated 15% greater precision than the econometric approach, attributed to ANN's capacity to automatically establish non-linear correlations between variables without human intervention. Similarly, Box-Jenkins was outperformed by ANNs and support vector machines in forecasting construction work gross values in Hong Kong using univariate datasets from 1983 to 2014 [16], [17].

2.2.2 Machine Learning Models

ML-based forecasting models have demonstrated improvements over traditional econometric methods by reducing human judgment in pattern evaluation. ANN exemplified this advantage by outperforming Box-Jenkins in forecasting construction work gross values in Hong Kong using 1983-2014 data [17]. While proving ML algorithms' superiority, the study was limited by using only univariate time series data without incorporating additional features.

Several years later, the study at [18] confirmed ANN's effectiveness for multivariate time series analysis. In forecasting housing sales in a Turkish city, ANN achieved an R^2 value of 0.94, surpassing Support Vector Machine, Linear Regression, Gaussian Process Regression, and Regression Tree models. The study employed seven economic indicators but omitted demographic features from housing demand analysis.

The authors of [19] conducted comprehensive research on residential construction demand forecasting in Jordan using 23 economic indicators collected from January 2007 to March 2022. Unlike [18], this study introduced a hybrid linear-based model (Elastic-Net) that achieved the highest accuracy among 11 algorithms evaluated, including ANN. Interest rates were identified as the dominant factor affecting Jordanian housing demand. However, the study focused exclusively on economic indicators, neglecting potentially beneficial demographic features.

2.3 Research Gaps

2.3.1 Dataset Limitations

Many housing demand forecasting studies are constrained by limited feature numbers. For instance, Lam and Oshodi [17] employed a univariate dataset (total real estate gross sales) to forecast Hong Kong housing demand. While single-feature datasets simplify processes and accommodate limited data sources, they risk overfitting and may inadequately reflect real-world scenarios where multiple factors interact complexly [20]. Univariate models frequently miss important correlations and patterns captured in multiple-input models.

Additionally, some studies exclude demographic indicators from predictive models. For example, Emec and Tekin [18] used economic features (interest rate, consumer confidence, consumer price, and construction confidence indexes) to forecast housing demand in Konya City, Turkey, but excluded demographic indicators such as population, migration, and age structure. Conversely, Hong [2] emphasized demographic features when addressing New Zealand's housing shortage.

2.3.2 ML Algorithm Limitations

It is noted in [6] that advanced algorithms like XGBoost and ANN present interpretability challenges due to their “black box” nature, particularly when explanations are required. High dimensionality and complex internal representations complicate clear mechanism interpretation for stakeholders unfamiliar with machine learning concepts. Developing methods for deeper interpretation of ANN operations represents a significant area for future research.

Furthermore, most studies employ basic ANNs not specifically designed for sequential time series data [21]. Basic ANNs (Feedforward Neural Networks) cannot effectively determine how preceding data influence subsequent points and final results. Advanced ANNs like RNNs and Long Short Term Memory networks (LSTM) address this limitation by incorporating memory or recurrent networks from previous inputs [21]. Implementing these advanced ANNs in housing demand forecasting could improve accuracy for time-series data analysis.

3 METHODOLOGY

This research applies machine learning algorithms to develop multiple forecasting models aimed at predicting housing demand. In addition, it leverages feature importance assessment techniques to interpret the influence of individual input features on the predicted demand. Figure 2 presents a flowchart illustrating the general steps involved in a typical machine learning project. The process includes seven key stages, forming a cyclical workflow in which steps may be revisited for refinement or additional data collection, depending on the results obtained during the testing and evaluation phase. The cycle begins with data collection and concludes with the development of a reliable predictive model and the insights derived from it.

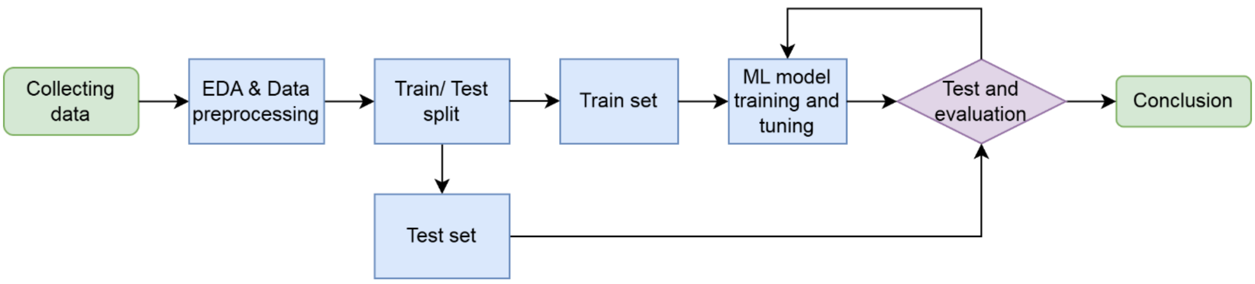

Figure 2: Overall machine learning workflow

3.1 Data Collection and Preprocessing

This study utilizes longitudinal data on New Zealand's demographic and economic indicators from the Reserve Bank of New Zealand database at https://www.rbnz.govt.nz/statistics/series/data-file-index-page. The dataset includes quarterly measurements from various macroeconomic indicators recorded primarily since the 1990s to September 2024 and published openly for the purpose of research and education. The database is categorized into separate series including exchange and interest rate, lending and monetary statistics, NZ debt securities, Household, and Economy indicators, etc. Each series contains different titles (or indicators) marked by a single code. From the complete database, we selected 16 indicators based on their demonstrated influence on construction industry in general and on residential housing demand in specific, as stated and utilized in previous research [22], [23] . Table 1 shows the indicators that were extracted from the database for the use of this research.

Table 1: Dataset features

No	Indicators	Code	Period	Unit	Peridiocity	Category
1	Population	M12	03/1991-09/2024	Quarterly	Number	Demography
2	Migration	M12	03/2000-09/2024	Quarterly	Number	Demography
3	Consumers Price Index (CPI)	M1	03/1991-09/2024	Index	Quarterly	Economy
4	Private consumption expenditure	M2	09/1995-09/2024	NZDm	Quarterly	Economy
5	Number of building consent (Target)		03/1993-09/2024	Number	Quarterly	Economy
6	Gross fixed capital formation - Residential buildings	M3	03/1993-09/2024	NZDm	Quarterly	Economy
7	Domestic trade - Retail	M4	12/1992-09/2024	NZDm	Quarterly	Economy
8	Domestic trade - Wholesale	M4	12/1992-09/2024	NZDm	Quarterly	Economy
9	GDP	M5	06/1987-06/2024	NZDm	Quarterly	Economy
10	National and household saving	M6	1972-2023	NZDm	Yearly	Economy
11	Import Volume	M8	03/1990-09/2024	NZDm	Quarterly	Economy
12	Export Volume	M8	03/1990-09/2024	NZDm	Quarterly	Economy
13	Unemployment rate	M9	03/1994-09/2024	%	Quarterly	Economy
14	House sales and price index	M10	03/1990-06/2024	Number	Quarterly	Economy
15	Mortgage interest rate	B20	02/1964-11/2024	%	Monthly	Economy
16	Exchange rate NZD/USD	B1	06/1973-11/2024	NZD/USD	Monthly	Economy

Table 1 shows that some indicators have data before 1995, but collection was inconsistent, with varying and missing values. To ensure consistency across all features, we selected quarterly data from September 1995 to September 2024, yielding 117 observations, using earlier data only for calculating lagged features where available. The final dataset comprises 15 independent variables including Population, Net Migration, Consumers Price Index (CPI), Private Consumption Expenditure, Gross Fixed Capital Formation - Residential Buildings, Domestic Trade (Retail and Wholesale), GDP, National and Household Savings, Import and Export Volumes, Unemployment Rate, House Sales and Price Index, Mortgage Interest Rate, and Exchange Rate NZD/USD. The target variable is the Number of Building Consents.

3.1.1 Exploratory Data Analysis (EDA)

Figure 3 presents a heat map [24] displaying the correlation coefficients among various economic indicators. The values in the matrix range from -1 to 1, representing both the strength and direction of linear relationships between pairs of variables. Notably, there is a strong positive correlation (0.84) between the *Number of Building Consents* and *Gross Fixed Capital Formation – Residential Building*, indicating that these two indicators tend to move in tandem. In contrast, a moderate negative correlation (-0.57) exists between building consents and the unemployment rate, suggesting that higher unemployment is associated with a decline in the number of building consents issued. Other variables exhibit weak or negligible correlations, such as the association between house sales to other households and household savings, which shows a low coefficient of 0.12, implying no significant linear relationship.

3.1.2 Data Cleaning

3.1.2.1 Missing data

There was no net migration data collected before 2000. Since the number of rows is quite limited, instead of removing those data points, mean value will be filled in those voids (approximately 6,406). This technique is simple but does not reflect the fluctuation of data within that period especially in the context of temporal dataset [25]. The study chose this option to reduce data preprocessing time. However, future studies can improve this using other advanced technique such as Denoising Autoencoder [26].

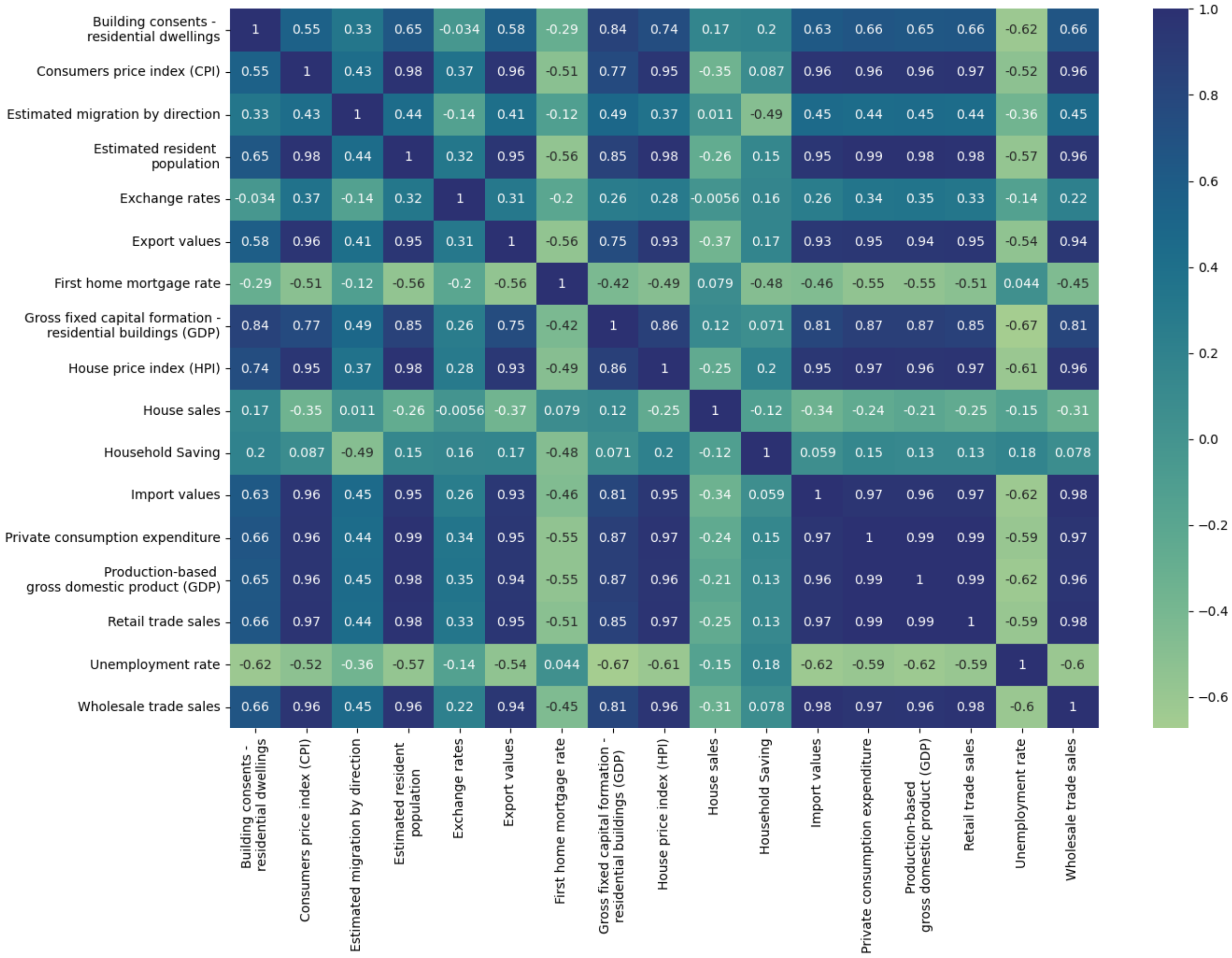

Figure 3: Feature correlation

3.1.2.2 Outliers

Outliers are observed in wholesale trade volumes, household savings, and residential housing consents. Figure 4 illustrates these three variables, all of which show a noticeable peak following the COVID-19 lockdown in 2021. The timing of these outliers aligns across variables, suggesting they are the result of a rare and significant event rather than data errors. One explanation, supported by several studies, is that during the pandemic, people tended to increase their savings as a precautionary response. However, following the rollout of vaccinations, many households began spending these accumulated savings, leading to the sharp peak observed in the household savings data [27].

To mitigate the effect of these outliers on prediction accuracy, while preserving their contribution as significant events, the original values were retained to reflect the impact of exceptional circumstances such as the Covid-19 pandemic. However, for linear ML algorithms including linear regression and Elastic Net, extreme values can substantially distort results. To address this challenge, the Winsorization technique [28] was implemented. Winsorization manages outliers by capping extreme values at specified percentiles (the 5^{th} and 95^{th} percentiles in this study) rather than completely removing them, thereby preserving their statistical significance while reducing their disproportionate influence.

3.1.3 Data Transformation

Each feature is scaled to a range between 0 and 1 by utilizing Min-Max normalization, which is calculated based on the feature's minimum and maximum values. This practice offers two main benefits: (1) it helps mitigate prediction distortion caused by extreme outliers, and (2) it reduces computational time, as the algorithms no longer need to process variables with differing scales and units. The mathematical formulation for Min-Max normalization is shown in Equation (5).

$$x_{norm} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{5}$$

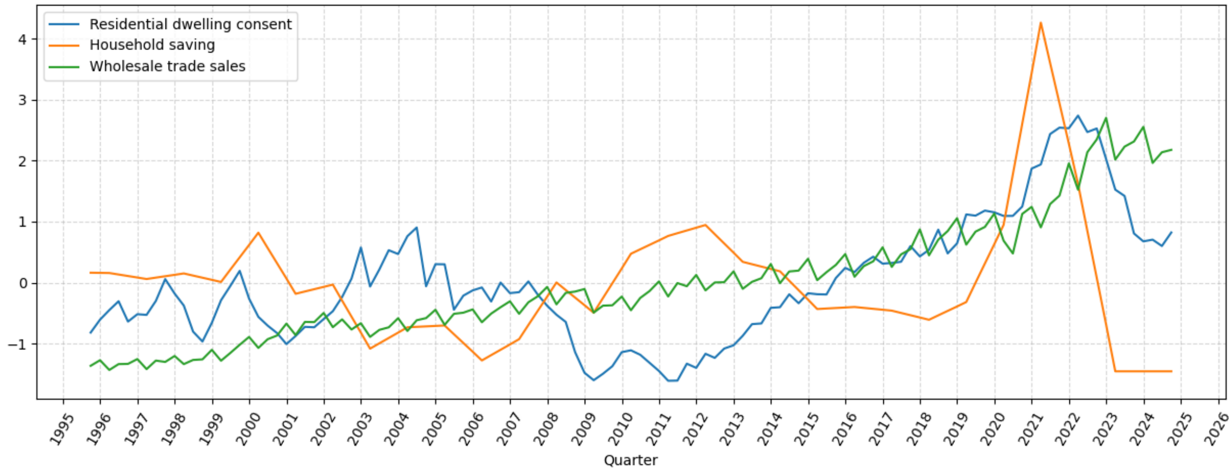

Figure 4: Outliers of 3 variables recorded at the same period

3.1.4 Feature Engineering

This study applies a feature engineering technique known as lagged features [29]. This technique is particularly useful for sequential time series datasets, as it divides the data into time frames. Each time frame contains a set of time steps used as input to predict the next value. The frame then shifts forward, and the process repeats, ensuring that previous data points are incorporated into the final predictive results. Since the dataset in this study is collected quarterly, lagged features are tested with values such as 2, 4, 8, and so on, until the model achieves the highest R^2 score. As such, this step also serves as part of the model tuning process.

	Building consents
Quarter	
1995-09-30	4927.0
1995-12-31	5413.0
1996-03-31	5773.0
1996-06-30	6105.0
1996-09-30	5340.0
...	...
2023-09-30	8653.0
2023-12-31	8355.0
2024-03-31	8420.0
2024-06-30	8182.0
2024-09-30	8690.0

	Building consents	Building consents_lag_1	Building consents_lag_2
Quarter			
1996-03-31	5773.0	5413.0	4927.0
1996-06-30	6105.0	5773.0	5413.0
1996-09-30	5340.0	6105.0	5773.0
1996-12-31	5617.0	5340.0	6105.0
1997-03-31	5592.0	5617.0	5340.0
...	...	...	...
2023-09-30	8653.0	10062.0	10300.0
2023-12-31	8355.0	8653.0	10062.0
2024-03-31	8420.0	8355.0	8653.0
2024-06-30	8182.0	8420.0	8355.0
2024-09-30	8690.0	8182.0	8420.0

Figure 5: Example of lagged features

This approach is beneficial for time series analysis because it incorporates temporal sequences into training and captures how past values influence future outcomes. It is particularly useful for models like linear regression, Elastic Net, and XGBoost, which do not inherently handle sequential data. In the context of economics, changes in an indicator may not have an immediate effect on the number of building consents. Instead, the impact may appear after a certain time lag. Lagged features are used here to reflect that delayed influence of one variable on another.

3.2 Model Development and Training

The research employed five ML algorithms: Linear Regression (baseline), Elastic Net, Ridge Regression, XGBoost, and RNN. The data was split in a 90:10 ratio for training and testing, with December 2021 serving as the split milestone. Figure 6 shows this partition, with the period before the milestone forming the training set and the remainder serving as the test set according to time sequence.

This 90:10 ratio, while not generally optimal for model evaluation, was selected due to the dataset's relatively small size. Allocating 90% for training ensured sufficient data points for effective learning. Additionally, the December 2021 milestone coincides with the peak of building consents before a significant decline, providing an ideal test case for the models' ability to predict substantial trend changes. This point also helps isolate the extreme effects of the COVID-19 event while testing the models' capability to identify the subsequent downward trend.

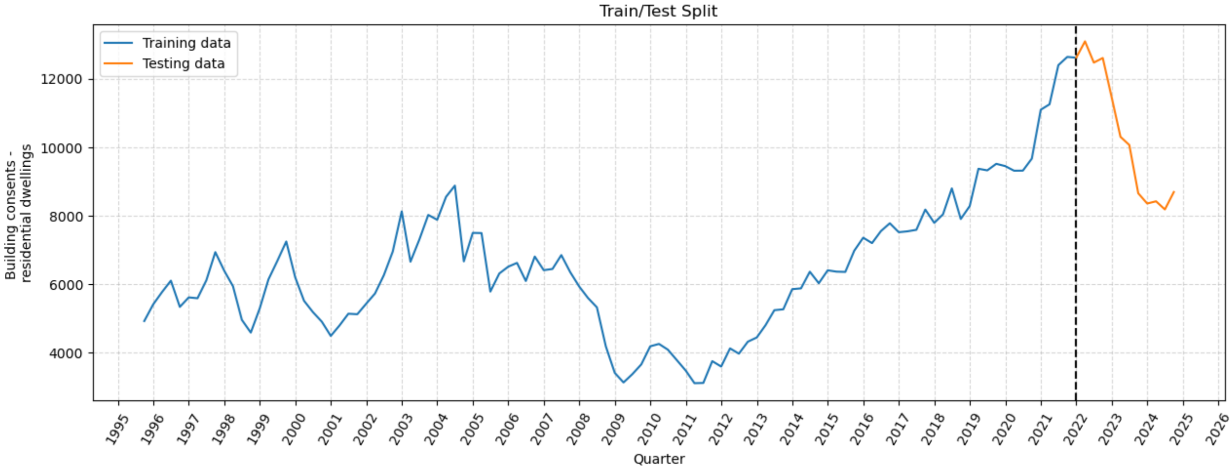

Figure 6: Train-test data split showing December 2021 milestone

For model validation, time-series cross-validation was implemented to prevent optimistically biased results common with traditional cross-validation methods applied to temporal data. As shown in Figure 7, unlike conventional cross-validation, time-series cross-validation ensures validation sets chronologically follow training sets, better simulating real-world forecasting conditions where future data remains unavailable during model training.

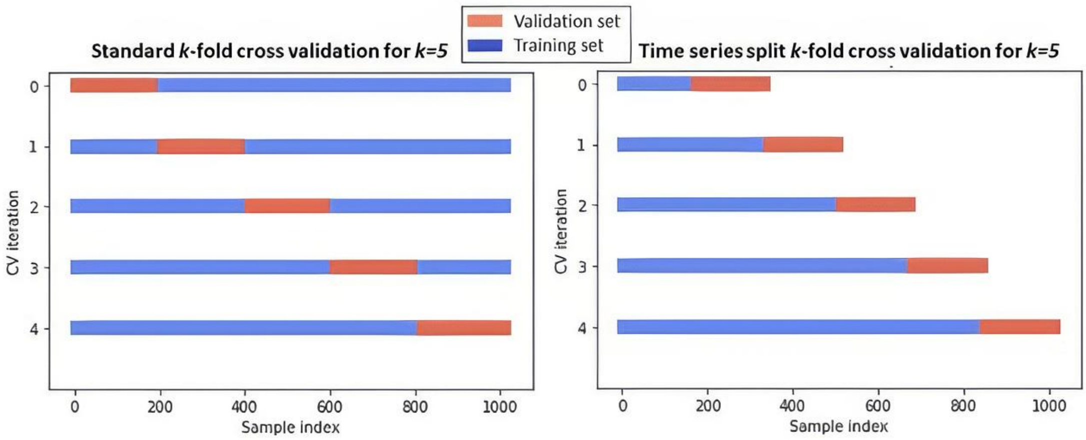

Figure 7: Traditional vs. time-series cross-validation approaches

3.2.1 *XGBoost Training Process*

The XGBoost model training proceeded through three distinct stages. Initially, the entire training set was fitted to a basic XGBoost model with randomly selected parameters to establish feature importance rankings. Next, the model was repeatedly trained with progressively reduced feature sets, with each iteration eliminating less important features based on their previously determined importance scores. Table 2 presents selected results from this iterative process, highlighting that optimal performance (R^2 = 75.26%) was achieved with 24 features at a threshold value of 0.004.

Table 2: XGBoost feature selection results (selected values)

Threshold	Number of Features (n)	R^2 (%)
0.000	48	14.23
0.000	42	58.70
0.004	27	-34.27
0.004	25	-14.28
0.004	**24**	**75.26**
0.005	23	39.67
0.006	19	54.12
0.066	4	-62.89
0.244	2	9.54

In the final stage, these 24 optimal features were used to train the refined model with hyperparameter tuning, including number of trees ($\text{n\_estimators} = 10{,}000$), maximum tree depth ($\text{max\_depth} = 7$), and learning rate ($\text{learning\_rate} = 0.3$). This hyperparameter optimization further improved the R^2 score from 0.75 to 0.77.

3.2.2 RNN Architecture and Training

The RNN model architecture, illustrated in Figure 8, is designed to predict housing demand using a structure with four layers: an input layer, two hidden layers, and an output layer. The input layer receives the data, which is processed by the first hidden layer (200 neurons) and the second hidden layer (100 neurons) to identify patterns. The output layer then produces a single prediction for the Number of Building Consents. To prepare the data for the RNN, we organized it into a 3D array with the shape [None, 1, 48]. “None” represents the number of quarters in the training set (flexible for any size), “1” means we use data from one quarter at a time, and “48” is the number of features. These 48 features come from 16 variables—15 economic and demographic indicators from Table 1, plus the Number of Building Consents—each including its current value and two previous values (lags). This results in $16 \times 3 = 48$ features, allowing the RNN to use past and present data to forecast future housing demand.

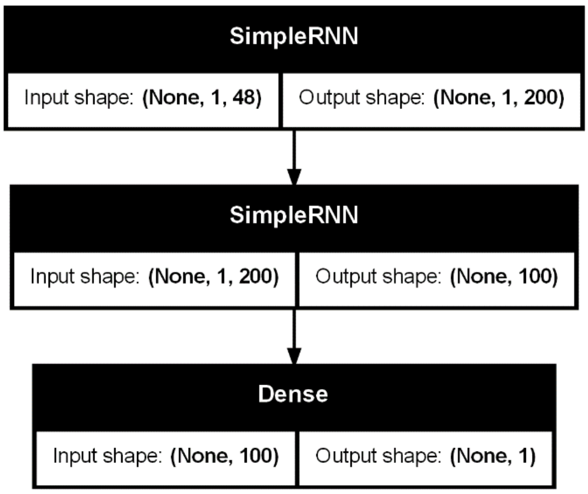

Figure 8: RNN model architecture

For the RNN model, the *tanh* activation function was selected for both hidden layers to capture non-linear relationships between variables and mitigate the vanishing gradient problem that commonly affects sequential data processing. The output layer employed a linear activation function, appropriate for regression tasks where predictions need not be bounded within a specific range. The model used the Adam optimizer [30], chosen for its computational efficiency and effectiveness with non-convex loss functions typical in neural networks.

3.2.3 Elastic Net and Ridge Regression Tuning

Hyperparameter tuning for the Elastic Net model explored multiple parameter combinations, including:

- Number of iterations: 1000, 1100, 1200
- Regularization strength (α): 0.01, 0.1, 0.5, 1.0, 5, 10.0
- L1/L2 ratio: 0, 0.01, 0.05, 0.1, 0.2, 0.4, 0.6, 0.8, 1
- Time series cross-validation splits: 4, 8, 12, 16, 24

Optimal performance (R^2 = 0.69, SMAPE = 12.95%) was achieved with 1100 iterations, α = 0.01, L1/L2 ratio = 0, and 24-fold cross-validation. Notably, the L1/L2 ratio of 0 effectively transformed the model into pure Ridge Regression, prompting a separate investigation of Ridge Regression with focused regularization strength values (α = 0.01, 0.05, 0.1, 0.5, 1). This Ridge Regression model achieved R^2 = 0.77 with α = 0.01, outperforming the initial Elastic Net implementation.

3.3 Evaluation Metrics

This study uses a set of complementary evaluation metrics to assess model accuracy from different angles. Each metric captures a specific aspect of prediction quality, allowing for a more complete understanding of model performance. Together, they address both absolute and relative accuracy, account for scale-dependent and scale-independent measures, and reflect varying sensitivity to error patterns.

3.3.1 Mean Absolute Error (MAE)

MAE [31] measures the average magnitude of errors without considering their direction, making it particularly useful for housing demand forecasting where the absolute scale of prediction errors directly impacts planning decisions. Lower MAE values indicate better model performance.

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{6}$$

where y_i is the actual value and $\hat{y}_i$ is the predicted value.

3.3.2 Mean Bias Error (MBE)

MBE [32] reveals systematic biases in predictions, indicating whether a model consistently over- or under-predicts housing demand:

$$\text{MBE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i) \tag{7}$$

A positive MBE indicates overprediction while negative values show underprediction. While MBE = 0 suggests no overall bias, errors may still be large but offset each other, necessitating analysis alongside other metrics.

3.3.3 Root Mean Squared Error (RMSE)

RMSE, widely used in forecasting literature [31], [33], provides an error measure in the same units as the original data:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{8}$$

RMSE emphasizes larger errors because of its squared term, making it ideal for housing demand forecasting where significant prediction errors can lead to major economic impacts. However, its sensitivity to outliers requires thorough preprocessing when extreme values occur.

3.3.4 Symmetric Mean Absolute Percentage Error (SMAPE)

SMAPE, initially defined by [34], expresses errors as percentages, making it easier to interpret and compare results across different datasets. This is especially useful when presenting findings to policymakers. Its value ranges from 0% to 200%, with 0% indicating a perfect fit and 200% representing extremely poor model performance. Equation (9) describes how SMAPE is calculated.

$$\mathrm{SMAPE} = \frac{100\%}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{\left(\frac{|y_i| + |\hat{y}_i|}{2}\right)} \tag{9}$$

3.3.5 Coefficient of Determination (R^2)

R^2 measures the proportion of variance in the dependent variable that can be explained by the independent variable(s) [35]. The R^2 value ranges from 0 to 1, with higher values indicating better model performance through greater explained variance. In social science research, an R^2 value from 0.51 to 0.99 is considered indicative of a high accuracy model [36]. R^2 is calculated as per Equation (10).

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{10}$$

where $\bar{y}$ is the mean of the actual values.

3.3.6 Mean Absolute Scaled Error (MASE)

MASE compares model performance against a naïve forecast (one-step persistence forecast) [37], making it particularly valuable for evaluating time-series models:

$$\mathrm{MASE} = \frac{\frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|}{\frac{1}{n-1}\sum_{i=2}^{n}|y_i - y_{i-1}|} \tag{11}$$

MASE < 1 indicates better performance compared to the naïve model, providing a scale-independent assessment of relative predictive ability. This is especially relevant for housing demand forecasting, where seasonal and cyclic patterns may be captured by simple persistence models.

4 RESULTS AND DISCUSSION

4.1 Model Performance Comparison

Table 3 presents the performance comparison across all five forecasting models using six complementary evaluation metrics.

Table 3: Model performance comparison

No.	Model	MAE	MBE	RSME	R^2	SMAPE (%)	MASE
1	Linear regression	1946.97	-1946.97	1977.29	-5.17	113.98	9.46
2	Elastic Net	479.62	-235.99	532.05	0.69	12.95	0.63
3	Ridge Regression	341.73	**70.11**	453.14	0.77	10.76	**0.45**
4	XGBoost	392.22	-117.91	436.35	0.77	11.48	0.97
5	RNN	**295.29**	170.62	**341.28**	**0.89**	**8.29**	0.88

The RNN model demonstrates better performance across all key indicators with the highest R^2, lowest SMAPE, lowest MAE, and lowest RMSE. This consistent phenomenon across different measurement approaches confirms RNN's suitability for housing demand forecasting in the New Zealand context. The model's positive bias (MBE = 170.62) indicates a tendency toward overprediction, which may be strategically advantageous from a housing policy perspective where slight overestimation carries fewer societal costs than underestimation.

Ridge Regression and XGBoost form a secondary performance tier with identical R^2 values but different error patterns. Ridge Regression achieves lower SMAPE and superior MASE, indicating better performance relative to both actual values and naïve forecasts. Conversely, XGBoost shows slightly better RMSE, suggesting more consistent error distribution. Their opposing bias directions—Ridge Regression's overprediction versus XGBoost's underprediction—reveal how models with similar overall accuracy can exhibit substantially different prediction characteristics.

Elastic Net follows with moderate performance, while Linear Regression demonstrates severe inadequacy with negative R^2 (-5.17) and very high error values across all metrics. The negative R^2 value indicates that Linear Regression performs substantially worse than simply predicting the mean value for all observations. This poor performance can be attributed to several factors: First, housing demand exhibits strong non-linear relationships with economic indicators that simple linear models cannot capture. Second, the presence of complex temporal dependencies in the dataset requires models that can account for sequential patterns. Third, multicollinearity among economic indicators likely creates instability in the coefficient estimates. Finally, the dataset contains several outliers, particularly following the COVID-19 pandemic, which disproportionately affect Linear Regression without regularization. These findings underscore the importance of either regularization techniques (as demonstrated by the much better performance of Ridge Regression) or advanced algorithms capable of modeling non-linear relationships for this application.

Regarding relative performance against naïve forecasting (measured by MASE), all models except Linear Regression achieved values below 1, confirming their superiority over simple persistence forecasts. Ridge Regression's particularly low MASE suggests it captures temporal patterns more than twice as effectively as a naïve approach, despite not being specifically designed for time-series data.

The superior performance of the RNN model indicates that the relationship between the input features and the target variable is more complex than what a simple linear model can capture. While models such as Elastic Net, Ridge Regression, and XGBoost perform well on general tabular datasets, they are not inherently designed to capture sequential patterns in time-series data unless enhanced through feature engineering. In contrast, RNNs are built with connected neurons that retain information from previous time steps and use it in subsequent computations. This architectural design enables RNNs to better capture temporal dependencies, making them more suitable for time-series forecasting tasks like this one.

4.2 Prediction Results

Figure 9, Figure 10, and Figure 11 illustrate the comparison between actual values and predictions made by Elastic Net/Ridge Regression, XGBoost, and RNN, respectively. These visualizations provide important insights into model behavior that complement the quantitative metrics presented in Table 3.

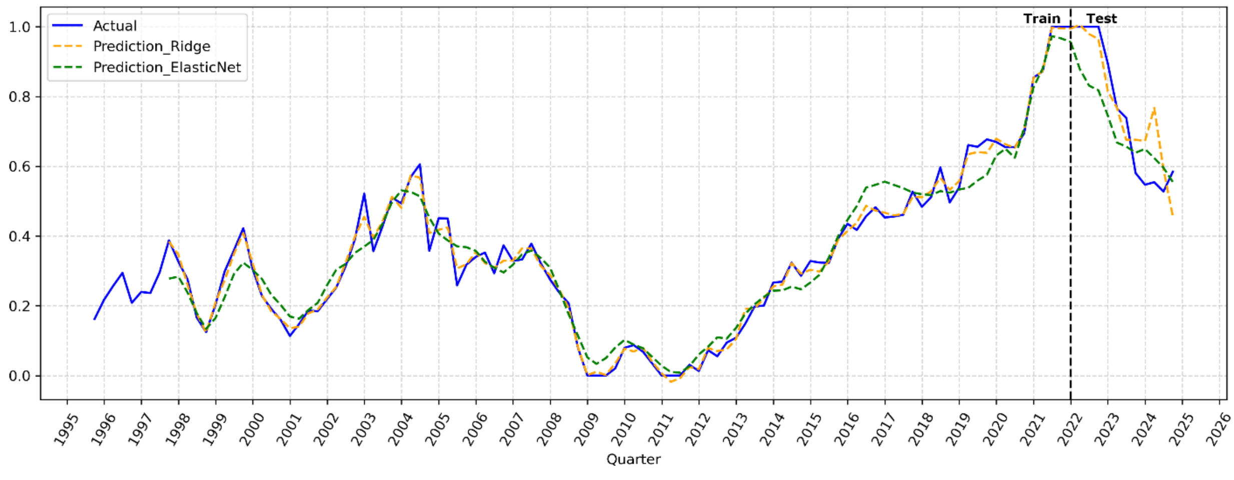

Figure 9: ElasticNet and Ridge Regression prediction result

The Elastic Net model shows notable discrepancies between actual and predicted values throughout both training and test periods. This is consistent with its lower R^2 score and higher error metrics. While it successfully captures the general upward trend from 2010 to 2021 and the subsequent downturn in the test set, the model frequently underestimates peaks and overestimates troughs, contributing to its negative MBE that indicates systematic underprediction.

Ridge Regression demonstrates improved tracking of actual values compared to Elastic Net, particularly during volatile periods (2000-2005 and 2015-2020). This improvement aligns with its superior performance metrics (MAE, SMAPE, and MASE). However, in the test set, Ridge Regression overestimates values in early 2022 before converging with actual trends by late 2023, reflecting its positive MBE.

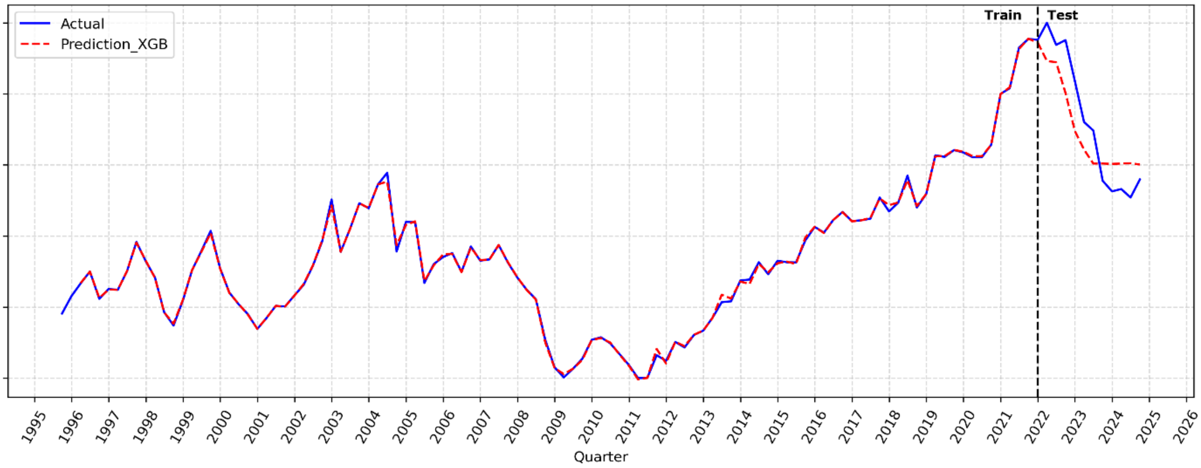

Figure 10: XGBoost prediction result

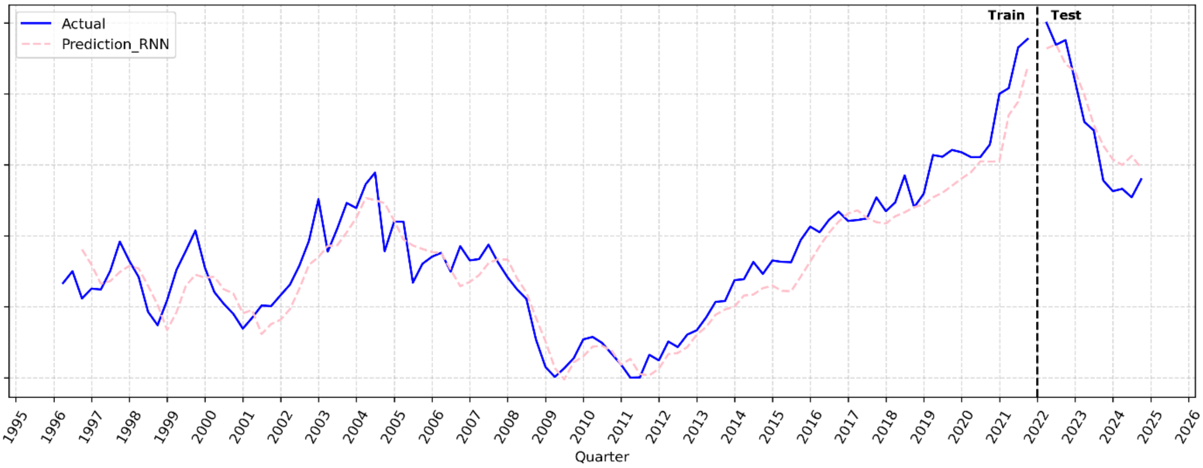

Figure 11: RNN prediction result

XGBoost exhibits remarkable alignment with the training data—the prediction line almost perfectly overlays the blue actual line throughout most of the training period. This suggests excellent in-sample fitting, which explains its competitive R^2 and RMSE values. However, this close training fit contrasts with its notable underprediction in the test set, particularly during the initial decline phase in early 2022, aligning with its negative MBE value.

The RNN model displays a distinct pattern compared to other models. During training, it shows smoother predictions with less reactivity to short-term fluctuations—evidenced by the pink dashed line's more gradual trajectory compared to the actual blue line's volatility. This suggests the model prioritizes capturing fundamental trends over fitting noise. Despite this apparent training bias, RNN achieves the best test set performance across all metrics, tracking the downward trend with remarkable accuracy while maintaining appropriate sensitivity to trend changes.

It is evident that while all models successfully capture the directional change in the test period, they differ significantly in their prediction characteristics. RNN's superior generalization despite higher training bias suggests it successfully learned the underlying temporal patterns rather than memorizing the training data. This balanced bias-variance tradeoff explains its consistently superior performance metrics and reinforces its suitability for housing demand forecasting applications.

4.3 Feature Importance Analysis

Figure 12 and Figure 13 present the feature importance rankings generated by Ridge Regression and

XGBoost, identifying key economic and demographic variables influencing housing demand in New Zealand. Despite methodological differences, both models highlight Consumers Price Index (CPI) and Gross Fixed Capital Formation – Residential Buildings as significant predictors. In XGBoost, CPI ranks highest (F-score = 79), followed by Gross Fixed Capital Formation and lagged CPI, indicating inflation's strong non-linear influence on demand. Ridge Regression, however, assigns the highest positive coefficient to Import Values (≈ 0.38), suggesting that increased imports correlate with higher housing demand, while Estimated Migration by Direction (lag 2) has a notable negative effect (≈ -0.22), and CPI shows a moderate negative impact (≈ -0.15).

These differences stem from the models' distinct approaches: Ridge Regression, a linear model, uses coefficients to capture magnitude and direction, potentially overemphasizing correlated features like Import Values and GDP. XGBoost, an ensemble of decision trees, captures non-linear relationships through data splitting, prioritizing features like CPI that frequently differentiate demand patterns. Other notable features include Unemployment Rate (positive coefficient in Ridge, moderate F-score in XGBoost), reflecting economic conditions' impact on housing, and lagged GDP in XGBoost, indicating broader economic trends. The limited influence of demographic variables like Population and Net Migration in both models suggests stable population growth in New Zealand may reduce their predictive power.

Figure 12: Ridge Regression feature importance assessment result

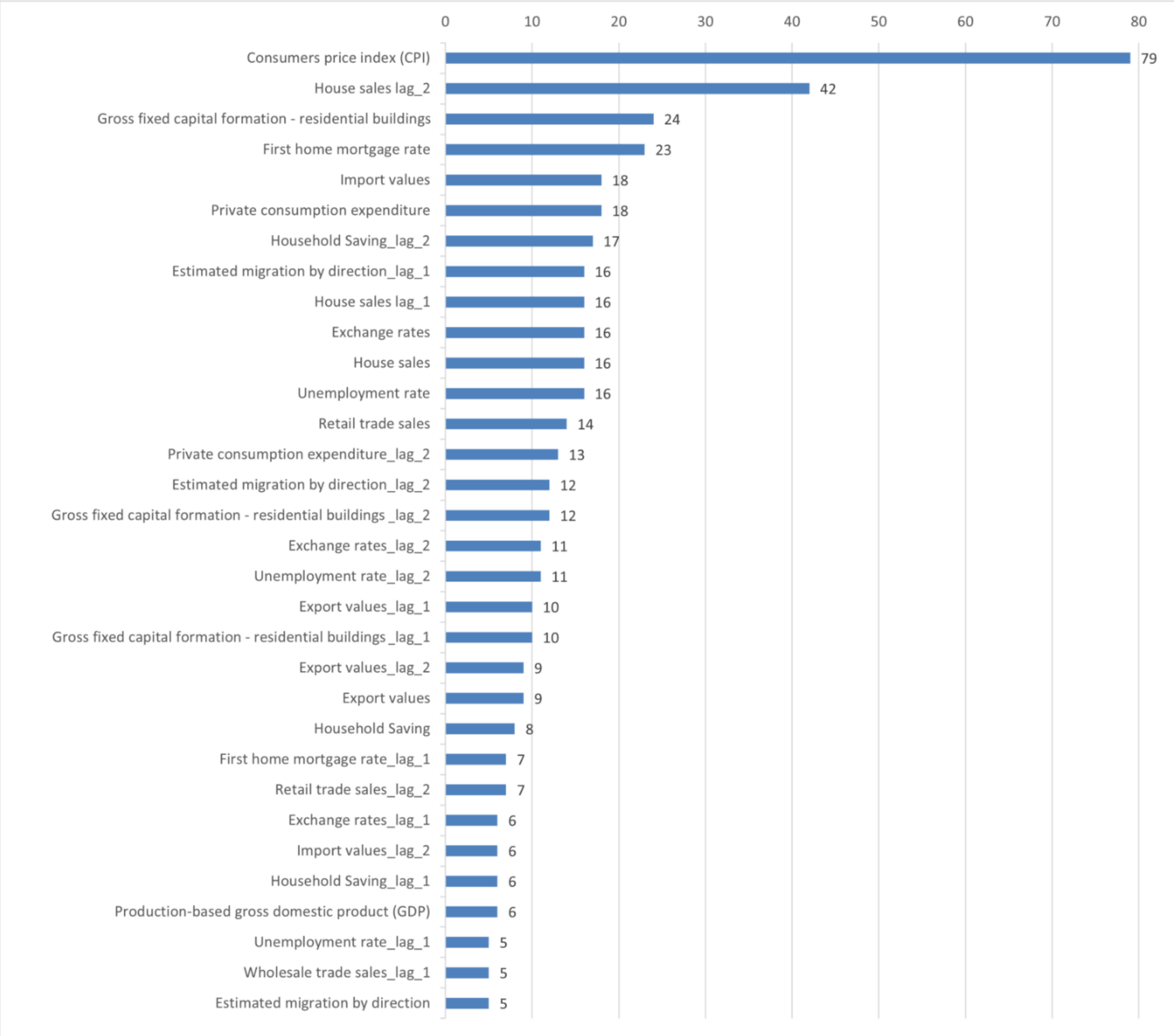

Figure 13: XGBoost feature importance assessment result

4.4 Summary of Findings

The dominance of the RNN model aligns with prior research on housing demand forecasting, such as [18] and [6]. However, studies like [19] and [38] show variation in model effectiveness, often influenced by differences in dataset size, feature sets, time frames, and observation frequency. These variations point to the localized nature of housing demand, which can differ significantly across countries and regions. Economic and demographic indicators may impact housing demand in different ways depending on the regional context, including factors such as economic conditions and population structure.

Feature importance analysis revealed diverse drivers of housing demand in New Zealand: economic factors like CPI, Gross Fixed Capital Formation – Residential Buildings, Import Values, and Unemployment Rate were prominent, reflecting the role of inflation, construction activity, trade, and economic conditions. Demographic variables (Population, Net Migration) showed limited impact, likely due to New Zealand’s stable population growth during the study period, lacking dramatic shifts like rapid growth or overpopulation seen elsewhere.

4.5 Limitations

While advanced models like RNNs achieve superior prediction accuracy, they present significant interpretability challenges. The "black box" nature of neural networks—with multiple hidden layers and complex interconnections—makes it difficult to assess feature importance, limiting causal insights. Similarly, feature importance assessment using Ridge Regression and XGBoost has limitations: Ridge coefficients may overemphasize correlated features such as Import Values and GDP, while XGBoost's F-scores lack directionality and are sensitive to dataset size.

Another limitation lies in the dataset's modest size (117 quarterly observations) that may affect long-term forecasting accuracy due to error accumulation, particularly when forecasts rely on prior predictions. Future research could integrate additional features such as seasonality indicators, long-term trends, or enhanced lag features to improve extended forecasting performance. Second, the quarterly frequency may limit the model's ability to capture rapid market fluctuations, potentially reducing responsiveness to sudden changes. More granular time intervals could improve detection of short-term market dynamics. Third, the national-level data provides a general overview but may not accurately represent conditions in specific cities or regions where economic and population dynamics vary considerably. Regional datasets would enable more targeted insights and support location-specific policy decisions.

5 CONCLUSION

This study investigated the relationships between economic and demographic indicators and housing demand in New Zealand, evaluating five ML algorithms for forecasting accuracy using six complementary metrics (R^2, SMAPE, MAE, RMSE, MBE, MASE). The RNN model outperformed others across most metrics, underscoring the importance of temporal modeling in capturing complex sequential relationships for housing demand forecasting. The feature importance analysis identified economic factors like CPI, construction investment, import values, and unemployment as key drivers, while demographic indicators had minimal impact due to New Zealand's stable population growth. The divergent rankings between linear (Ridge Regression) and tree-based (XGBoost) models highlight the value of multiple approaches in understanding complex economic relationships. These findings address a critical research gap in housing demand forecasting in New Zealand, offering policymakers valuable tools to tackle the country's housing crisis—described as "among the most unaffordable" markets globally due to supply shortages [39]. Insights into key drivers like inflation and import values can inform strategies to balance housing supply and demand. Future research should explore expanding dataset dimensions, developing hybrid models, enhancing neural network interpretability, and integrating advanced error analysis to better understand model behavior, ultimately providing more actionable insights for addressing New Zealand's housing challenges.

REFERENCES

[1] I. Ministry of Business and Employment, "Building and Construction Sector Trends: Annual Report 2023," Ministry of Business, Innovation and Employment, 2023.

[2] L. Hong, "The need to build: The demographic drivers of housing demand," *N. Z. Initiat.*, 2021.

[3] H. M. A. A. Obaid, "Factors determining housing demand in Saudi Arabia," *Int. J. Econ. Financ. Issues*, 2020.

[4] StatsNZ, "Housing in Aotearoa: 2020," *www.stats.govt.nz*, 2020.

[5] H. Jiang, "Econometric techniques for estimating construction demand in Australia," in *Econometric techniques for estimating construction demand in Australia*, 2013. [Online]. Available: https://api.semanticscholar.org/CorpusID:150482008

[6] B.-H. Goh, "Forecasting residential construction demand in Singapore a comparative study of the accuracy of time series, regression and artificial neural network techniques," *Sch. Build. Real Estate Natl. Univ. Singap.*, 1998.

[7] I. El Naqa and M. J. Murphy, *What is machine learning?* Springer, 2015.

[8] T. Jiang, J. L. Gradus, and A. J. Rosellini, "Supervised machine learning: a brief primer," *Behav. Ther.*, vol. 51, no. 5, pp. 675–687, 2020.

[9] C. Ter Braak and C. Looman, "Regression," in *Data analysis in community and landscape ecology*, Cambridge University Press, 1995, pp. 29–77.

[10] H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc. Ser. B Stat. Methodol.*, vol. 67, no. 2, pp. 301–320, 2005.

[11] T. Chen and C. Guestrin, "Xgboost: A scalable tree boosting system," in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 2016, pp. 785–794.

[12] S. Esmailizade, A. Ebrahimi, H. Soltani, A. Sam, and M. Rahimi, "Machine Learning Approaches for Retail Forecasting: A Study on XGBoost and Time-Series Models," *Authorea Prepr.*, 2024.

[13] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent advances in recurrent neural networks," *ArXiv Prepr. ArXiv180101078*, 2017.

[14] S. A. Kader, N. M. Zayed, M. Faisal-E-Alam, M. Salah Uddin, V. Nitsenko, and Y. Klius, "Factors affecting demand and supply in the housing market: A study on three major cities in Turkey," *Computation*, vol. 10, no. 11, p. 196, 2022.

[15] S. Turgay, C. Kanbolat, and E. S. Kara, "Time Series Analysis of Housing Demand: A Forecasting Model for Ankara, Turkey," *Environment*, vol. 14, p. 15, 2022.

[16] K. C. Lam, O. S. Oshodi, and E. W. M. Lee, "Forecasting construction demand: A comparison of Box-Jenkins and Support Vector machine model," *9th Cidb*, p. 14, 2016.

[17] K. C. Lam and O. S. Oshodi, "Forecasting construction output: a comparison of artificial neural network and Box-Jenkins model," *Emeral Publ. Ltd.*, 2016.

[18] Ş. Emec and D. Tekin, "Housing demand forecasting with machine learning methods," *Erzincan Univ. J. Sci. Technol.*, 2022.

[19] F. Sammour, H. Alkailani, G. J. Sweis, R. J. Sweis, W. Maaitah, and A. Alashkar, "Forecasting demand in the residential construction industry using machine learning algorithms in Jordan," *Emeral Publ. Ltd.*, 2023.

[20] D. Jackson, R. Riley, and I. R. White, "Multivariate meta-analysis: potential and promise," *Stat. Med.*, vol. 30, no. 20, pp. 2481–2498, 2011.

[21] Y. Hua, Z. Zhao, R. Li, X. Chen, Z. Liu, and H. Zhang, "Deep learning with long short-term memory for time series prediction," *IEEE Commun. Mag.*, vol. 57, no. 6, pp. 114–119, 2019.

[22] R. O. Asamoah, B. K. Baiden, G. Nani, and E. Kissi, "Review of Exogenous Economic Indicators Influencing Construction Industry," 2019.

[23] G. B. Hua, "Modeling sectoral construction demand and its relationship with economic indicators," *Int. J. Constr. Educ. Res.*, vol. 8, no. 3, pp. 223–240, 2012.

[24] N. Iliinsky and J. Steele, *Designing data visualizations: Representing informational Relationships*. O'Reilly Media, Inc., 2011.

[25] A. Tawakuli, B. Havers, V. Gulisano, D. Kaiser, and T. Engel, "Survey: time-series data preprocessing: a survey and an empirical analysis," *J. Eng. Res.*, 2024.

[26] S. Ryu, M. Kim, and H. Kim, "Denoising autoencoder-based missing value imputation for smart meters," *IEEE Access*, vol. 8, pp. 40656–40666, 2020.

[27] H. Ren and Y. Zheng, "COVID-19 vaccination and household savings: An economic recovery channel," *Finance Res. Lett.*, vol. 54, p. 103711, 2023.

[28] J. H. Sullivan, M. Warkentin, and L. Wallace, "So many ways for assessing outliers: What really works and does it matter?," *J. Bus. Res.*, vol. 132, pp. 530–543, 2021.

[29] R. F. Engle and S. Kozicki, "Testing for common features," *J. Bus. Econ. Stat.*, vol. 11, no. 4, pp. 369–380, 1993.

[30] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *ArXiv Prepr. ArXiv14126980*, 2014.

[31] T. Chai and R. R. Draxler, "Root mean square error (RMSE) or mean absolute error (MAE)?–Arguments against avoiding RMSE in the literature," *Geosci. Model Dev.*, vol. 7, no. 3, pp. 1247–1250, 2014.

[32] J. E. Hunter and F. L. Schmidt, *Methods of meta-analysis: Correcting error and bias in research findings*. Sage, 2004.

[33] T. O. Hodson, "Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not," *Geosci. Model Dev. Discuss.*, vol. 2022, pp. 1–10, 2022.

[34] J. S. Armstrong, *Long-Range Forecasting: From Crystal Ball to Computer*, 2nd ed. Wiley-Interscience, 1985.

[35] S. Wright, "Correlation and causation," *J. Agric. Res.*, vol. 20, no. 7, p. 557, 1921.

[36] P. K. Ozili, "The acceptable R-square in empirical modelling for social science research," in *Social research methodology and publishing results: A guide to non-native English speakers*, IGI global, 2023, pp. 134–143.

[37] P. H. Franses, "A note on the mean absolute scaled error," *Int. J. Forecast.*, vol. 32, no. 1, pp. 20–22, 2016.

[38] B. Chen and J. Li, "Analysis and Forecast of Housing Demand Impact in China's Underdeveloped Areas Based on XGBOOST Algorithm," *BCP Bus. Manag.*, 2022.

[39] H. C. Bishop, "Ministry of Housing and Urban Development," in *Office of the Minister of Housing, Infrastructure and RMA Reform*, 2024.

DỰ BÁO NHU CẦU NHÀ Ở NEW ZEALAND: SO SÁNH CÁC MÔ HÌNH ELASTIC NET, XGBOOST VÀ RNN

Tóm tắt. Dự báo nhu cầu nhà ở là một chủ đề nghiên cứu phổ biến trên toàn cầu, với các công trình chủ yếu sử dụng các phương pháp kinh tế lượng truyền thống. Ứng dụng của máy học trong lĩnh vực này vẫn còn hạn chế, đặc biệt là trong bối cảnh New Zealand. Nghiên cứu này giải quyết vấn đề đó bằng cách triển khai các mô hình Elastic Net, XGBoost và Recurrent Neural Network để dự đoán nhu cầu nhà ở dân dụng tại New Zealand bằng cách sử dụng dữ liệu kinh tế và nhân khẩu học từ năm 1995 đến hiện nay. Các mô hình được đánh giá bằng cách sử dụng 6 chỉ số hiệu năng (R^2, SMAPE, MAE, RMSE, MBE và MASE), với mô hình RNN đạt được độ chính xác cao nhất. Kết quả thí nghiệm chứng minh rằng các thuật toán học máy cải thiện đáng kể dự báo nhu cầu nhà ở với các mô hình chuỗi thời gian vượt trội hơn các phương pháp tiếp cận truyền thống. Kết quả phân tích tầm quan trọng của các thuộc tính đã xác định chỉ số giá tiêu dùng, đầu tư cho xây dựng, giá trị nhập khẩu và thất nghiệp là các động lực chính, trong khi các yếu tố nhân khẩu học cho thấy tác động hạn chế trong việc thúc đẩy nhu cầu nhà ở. Những phát hiện này cung cấp thông tin có giá trị cho các nhà hoạch định chính sách và các công ty xây dựng giải quyết các thách thức về nhà ở của New Zealand. Các nghiên cứu trong tương lai nên tập trung vào việc mở rộng kích thước tập dữ liệu và triển khai các kỹ thuật tối ưu hóa cũng như cung cấp các phương cách diễn giải tiên tiến hơn để tinh chỉnh các mô hình này.

Từ khóa. Máy học, Mạng nơ-ron, Nhu cầu nhà ở, Mô hình dự báo, Phân tích chuỗi thời gian.

Received on Arpil 14 - 2025

Revised on May 03 - 2025