2023 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE) | 979-8-3503-4107-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/CSDE59766.2023.10487658

# Real Estate Price Prediction on GenerativeLanguage Models

Yun Zhao
School of IT & Systems
Faculty of SciTech
University of Canberra
Australia

Girja Chetty
School of IT & Systems
Faculty of SciTech
University of Canberra
Australia
girija.chetty@canberra.edu.au

Dat Tran
School of IT & Systems
Faculty of SciTech
University of Canberra
Australia

***Abstract*— Real estate prediction is an important field of study that can assist home sellers and property investors in making informed decisions to maximize their profits. However, predicting real estate prices presents a significant challenge in the temporal dimension because house prices fluctuate over time in response to market dynamics and are influenced by a complex array of factors, such as location, quality of the house and so on. This paper focuses on leveraging the power of transformer-based language models and self-attentions for real estate house price prediction. The Transformer architecture is well-known for its ability to understand the relationships between words or tokens in a sequence of human language and process the entire sequence in parallel, enabling more efficient and scalable computations. Our study explores the fine-tuning of attention mechanisms and output hidden states from Transformer-based models, comparing their performance against baseline models. Through our experimentation and analysis, the results demonstrate that the transformer-based attention models outperform the baseline models for real estate price prediction. We also discovered that utilizing self-attentions from unsupervised text learning can enhance the accuracy of real estate price prediction.**

***Keywords*— real estate, house price prediction, transformer attention, generative AI**

## I. Introduction

The real estate industry plays a crucial role in the economic growth and individual wealth for most countries. However, the real estate market is a complex and dynamic environment that is heavily influenced by various factors, such as economic cycles, interest rates, government policies, and property supply and demand. Understanding and navigating these factors can be challenging for humans. Machine learning techniques have been extensively chosen for the study of real estate price prediction for many years in both academic and real estate industry, primarily due to their capability to effectively handle complex environmental data and extract valuable insights, patterns, and relationships from extensive datasets. In traditional machine learning algorithms for real estate price prediction, the focus is typically on handling baseline real estate attributes such as location, bedrooms, bathrooms, and suburbs to analyse the relationships among these factors. However, these traditional machine learning architectures usually ignore the analysis of other critical factors, for instance, they may not consider the quality of the property, the surrounding environment of the suburb, or understand the supply and demand dynamics within the real estate market. Additionally, valuable insights from external sources, such as blogs from real estate experts, serve as important data for predicting real estate prices. Incorporating these essential pieces of information into traditional baseline machine learning algorithms for analysis can be a challenging task. Recently, with the growing popularity and power of large language models, there is now an opportunity to include that above important information about real estate market into large language models during comprehensive real estate price prediction analysis. Large language models possess the ability to understand the relationships between words or tokens in a sequence of human language, often referred to as "contextualized representations" or "attention weights". This capability enables the models to perform real estate price prediction from a natural language understanding task point of view, like how humans would approach the task. Our study focuses on investigating the performance of transformer-based models for real estate price prediction. We conduct experiments by fine-tuning the pre-trained parameters while incorporating an additional neural network to modify attention weights. Additionally, we explore the benefits of applying in-context unsupervised learning prior to fine-tuning the model. Through these experiments and developments, this study aims to enhance the accuracy and effectiveness of real estate price prediction by leveraging the capabilities of pre-trained large language models.

In summary, the contributions of this paper are: (1) Development of a novel approach for using language model within data augmentation by adding extra information for real estate price prediction (2) Investigating the performance of transformer-based models for real estate price prediction (3) Providing an example to leverage large language models' ability to compare with the limitations of traditional machine learning algorithms for predicting real estate price.

The paper is structured as follows: The next section presents an overview of the relevant literature on house price prediction, including standard machine learning models, deep learning-based models, and transformer models for Natural Language Processing (NLP). Section 3 introduces the modified pre-trained model and provides details on the experimental evaluation. Finally, Section 4 concludes the paper by presenting the findings of this study and outlining future research plans.

## II. Related work

### A. *Real Estate Price Prediction*

Real estate price prediction has been the subject of study by both the academic and real estate industry using various techniques, such as empirical, numerical, statistical, and machine learning approaches. In the paper [1], the author Bragoudakis et al. point out that Error-Correction Model (ECM) or Vector Error-Correction Model (VECM) are common approach when employing empirical models for real estate price prediction, which captures changes in house prices based on explanatory variables. Another study conducted by Hjort [2] focuses on real estate price prediction in the Norwegian housing market using statistical models - Automated Valuation Models (AVMs). He compares different methods for uncertainty quantification, including split conformal inference and conformalised quantile regression. His finding suggests that the methods employing conformalised quantile regression produce narrower confidence regions compared to split conformal inference. For machine learning approaches, recently study conducted by Wang et al. [3] developed a robust deep learning model for house price prediction, integrating heterogeneous data analysis and a joint self-attention mechanism to improve prediction accuracy. The model adopts the spatial transformer network (STN) techniques to process satellite maps and extract image features rather than using CNNs to achieve rotation invariance of satellite images. Additionally, a joint self-attention mechanism is introduced to capture combined relationships between public facilities data and house transaction data. This study compared the proposed model with other baseline machine learning algorithms and demonstrates that the proposed model achieves a low prediction error and outperforms the other models. The author Zaki et al. [4] conducted a study on house price prediction using a combination of XGBoost algorithm with outlier sum-statistic (OS) approach. The aim of the research was to explore the potential of machine learning algorithms, specifically the XGBoost algorithm, in improving the accuracy of house price prediction. The study compared the performance of the XGBoost algorithm with the traditional hedonic regression model, using 13 variables as inputs to predict house prices. The results showed that the XGBoost algorithm achieved a significantly higher accuracy of 84.1%, while the hedonic regression model had a much lower accuracy of 42%. The research concluded that the integration of machine learning techniques, particularly XGBoost, can provide practical and accurate predictions of house prices, which can be beneficial for property sellers and buyers in the real estate industry. Ozogur et al. [5] proposed a hybrid model that combines various techniques including linear regression, clustering analysis, nearest neighbour classification, and Support Vector Regression (SVR) and then using the output of one method as the input of another method for house prices estimation. The proposed model is trained using housing data collected from the Kadikoy district in Istanbul and evaluated against the KAGGLE house dataset, and the result shows that the proposed hybrid model yield superior results compared to several other regression methods. Another study conducted by author Temur et la. [6] presented a hybrid model by employing ARIMA (Auto Regressive Integrated Moving Average) as a linear model, LSTM (Long Short-Term Memory) as a nonlinear model and combining LSTM and ARIMA to estimate the time series of housing sales on real estate sales forecasting in Turkey. Their proposed hybrid model produced the best performance among these three models and demonstrating the lowest error rate

### B. *Transformer and Attentions*

The Vaswani et al. [7] introduced "Attention Is All You Need" as a machine learning approach based on the transformer architecture they developed. This model is solely relying on attention mechanisms by leveraging multi-headed self-attention to achieve remarkable performance in machine translation tasks. Another comprehensive research conducted by Radford et al. [8] demonstrated that training on large language models with extensive and diverse datasets can enable models to acquire the ability to perform diverse tasks without the requirement of supervision. Their findings suggest that large language models hold promise for building language processing systems that can learn from naturally occurring demonstrations and perform tasks effectively. Pfeiffer et al. [9] developed a novel framework called AdapterHub, which significantly improve the process of training pre-trained large language models for specific tasks. This framework builds upon the widely used transformer architecture, enhancing its capabilities by seamlessly integrating small neural modules known as adapters. By adding these adapters to existing transformer models, this framework facilitates efficient and straightforward transfer learning across various languages related tasks. According to the articles [10-16], these papers discuss in-context learning of language models and provided various perspectives and evidence regarding the explainability of language models. The feature of explainability in language models might hold the potential to improve the performance of real estate price prediction by using natural language-based explanations of sample data.

## III. TRANSFORMER BASED MODEL FOR HOUSE PRICE PREDICTION

In this section, we present our model of experiment for house price predication based on generative pre-trained transformer. This experiment is an extension of my previous research study [17, 20] that examines use of deep learning combining with eXtreme Gradient Boosting (XGBoost) for real estate appraisal, by analysing historical sale records together with visual content, with online house pictures, and by scoring each image from an aesthetic point of view to make a house price prediction. The experiment shows that an improvement in performance of house price prediction accuracy, with replacing the last output layer with XGBoost. We will discuss the key components of the new model we built using generative pre-trained transformer. However, before getting into that, we will first explain why we chose a language model for real estate price prediction. Following that, we will introduce data preparation and cover some of the notations we used in our experiment..

Identify applicable funding agency here. If none, delete this text box.

### A. *Why Choosing Language Model ?*

The main goal of choosing a language model for this experiment is that certain factors influencing real estate prices are difficult to represent using basic statistical information such as number of bedrooms, bathrooms and so on. Textual explanatory descriptions contain valuable information that are often hardly represented or captured with just a single numerical value, and this information might play a crucial role in real estate price prediction. For example, using textual descriptions, we can describe details about the quality of a house, its aesthetics, the surrounding environment, or even subjective feelings expressed by professionals. We think that large language models might have the ability to capture such complex and important information, enabling better predictions. For this study, we use the pre-trained GPT-2, a large language model with huge parameters, to predict real estate prices. Based on [8] the model was trained on a vast dataset of web pages, where its objective was to predict the next word based on the preceding words in a text. Since the dataset comes from different domains, which allow the model to encounter various natural language tasks. By employing the pre-trained GPT-2 mode, we aimed to explore its capability in comprehending textual explanatory descriptions during the prediction of real estate prices.

### B. *Data Preparation*

The experiment data was collected for past sales in city of Canberra, Australia from online Real Estate website called “Allhomes” (www.allhomes.com.au ). The data was separated as two datasets, dataset X collected from January 2019 to April 2019, and the other dataset X'was collected from November 2021 to July 2022. For dataset X each record contains address, bedrooms, bathrooms, ensuites, garages, carports, land size, unimproved value as shown in Table I, while dataset X' contains basic features like X but with property explanatory description context. We are particularly interested in investigating unsupervised property in-context learning approach based on dataset X' , whether it can improve house price prediction of dataset X.

TABLE I.
SAMPLE OF PROPERTY ATTRIBUTES

| *Address* | *Suburb* | *Bedrooms* | *Bathrooms* | *Garages* |
|---|---|---|---|---|
| *Pandanus Street, Fisher ACT 2611 | Belconnen | 3 | 1 | 1 |

Our dataset can be denoted as $[(x_1, y_1), \ldots, (x_n, y_n)]$, where $n$ is total samples of properties, and $x_n$ refers to basic features of the $n$-$th$ sample and $y_n$ refers to the corresponding house price of the $n$-$th$ property. The features of $x$ can be classified into three different groups in our dataset, which are the numerical data, such as bedrooms, bathrooms, carparks and so on, the categories data, such as suburb, property type, and textual explanatory descriptions, therefore, the dataset can be denoted by $x = [x_{num}, x_{cate}, x_{text}]$. As explained before the dataset $x$ was divided into two different timeframes and can be denoted as $x = [X, X']$ , where the sub-dataset $X$ doesn’t have property explanatory description context, and can be donated as $X = [X_{num}, X_{cate}]$, and the sub-dataset $X'$ can be donated as $X' = [X'_{num}, X'_{cate}, X'_{context}]$.

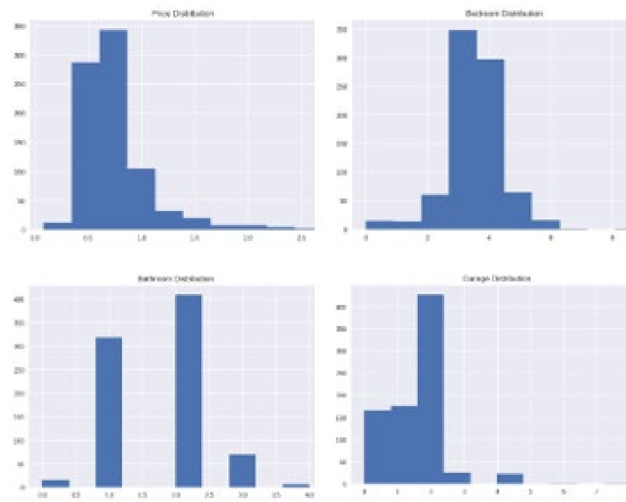

Fig. 1. Price, Bedrooms, Batheooms and Garage Data Distribution

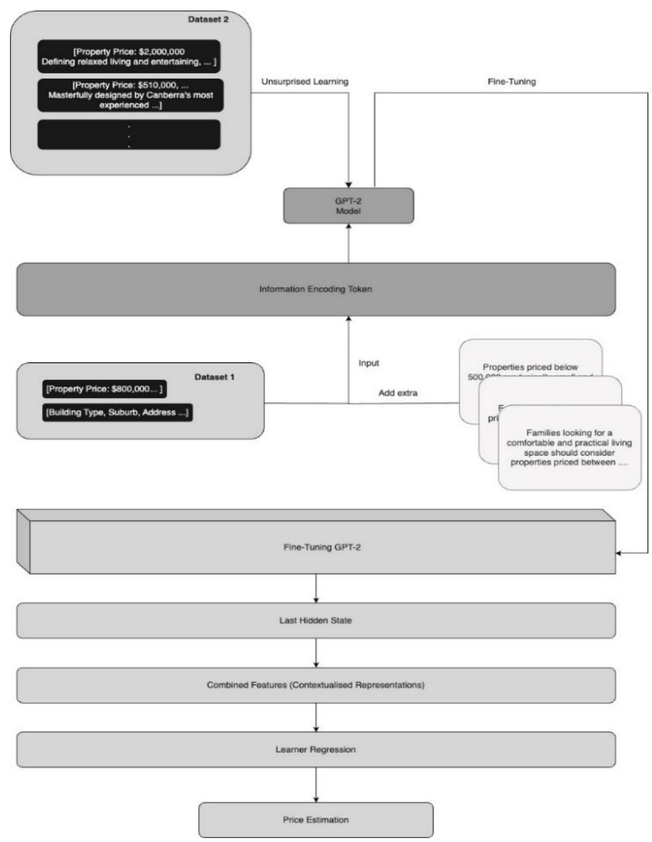


Fig. 2. The modified pre-trained GPT-2 framework for house price prediction experiment

For this experiment, we collected 1023 records (dataset $X$) and split 818 records to fine tuning the GPT-2 model and 205 records for testing. The Fig.1 shows general data distributions about prices, bedrooms, bathrooms, and garages. You can see that there are a few records with high values for price, bedrooms, bathrooms, and garages. However, most records are within the price range of $600,000 to $1,000,000, with 3 to 4 bedrooms, 2 bathrooms, and 2 garages. As a result, this dataset X contains many imbalanced data categories from prices, bedrooms, bathrooms, and garages. One approach, as highlighted by the authors in [18], is to apply SMOTE (Synthetic Minority Over-sampling Technique) for data augmentation, which helps address the class imbalance problem in machine learning. For this study we added extra information to the original data to enhance the model prediction without changing total number of samples during training. (See Fig. 2).

## C. *Generative Pre-Trained Transformer.*

Some of the earlier work based on Generative Pre-trained Transformer (GPT) involves using a Transformer architecture for the language modeling process [7,19]. The proposed Transformer architecture in this research adapts this model, so to allow effective learning of long-term dependencies and contextual information within the input text. GPT model has two key features for dealing with the language model tasks. The first feature is called generative pre-training, which is unsupervised training process on a very large corpus of text. This purpose of this process allows the GPT model to learn general language patterns and relationships, as well as develop a representation of language model that can be used for a variety of tasks. The second feature is called fine-tuning GPT-2 model network parameters with new training data. For this study, we are using GPT-2 pre-trained model involving unsupervised learning process on property explanatory descriptions with the dataset X' which has 1,000 samples to create new model M'. Then we take the new model M' and fine-tune it using supervised learning process on a real estate price prediction task. The goal of this experiment was to use the fine-tuning process to adapt the model's general language understanding capabilities, obtained through the previous unsupervised training process, for continued supervised training on a specific task to achieve low prediction error rate.

In this study, the unsupervised corpus of tokens, $D$, refers to a collection of property in-context description and basic attributes, where each token $D_i$ corresponds to a word of token in the property in-context description. The unsupervised corpus can represent $D_i = [d_1, \dots, d_i]$ , the objective of the unsupervised training process is to maximise the likelihood of predicting each token in the corpus of tokens $D_i$ based on the sequence of given the preceding context. The following equation shows the function for the unsupervised corpus of tokens $D_i$:

$$L(D_i) = \sum_{1}^{n} P(d_i \mid d_1, \dots, d_n ; \theta) \quad (1)$$

where $P$ represents the probability of predicting token $d_i$, given the preceding context $d_1, \dots, d_n$ and the model's parameters $\theta$.

The following provides an example of in-context unsupervised learning:

*Context: As a real estate agent, it is your responsibility to estimate the value of a property accurately. Based on the following information, you need to formulate an explanatory description that ensures an accurate house price prediction.*

*Basic Information:*

*'Category: A small residential property with shared public infrastructure in an affordable suburb of Crace'*
*'Price Range: Below $500,000'*
*'Suburb: Crace'*
*'Address:187 Bettong Ave, Crace ACT 2918'*
*'Bathrooms: 2:*
*'Bedrooms: 4'*
*'Garage: 2'*
*'Landsize: 400'*
*'Latitude: -35.20107 '*
*'Longitude: 149.0966118 '*

*Example of explanatory description for price less than $500,000:*

*"The property is located in the suburb of Crace, an affordable suburb with shared public infrastructure. It falls under the category of a small residential property. The price range for this property is below $500,000. The address is 187 Bettong Ave, Crace ACT 2918. It has 2 bathrooms, 4 bedrooms, and a 2-car garage. The landsize is approximately 400 square units. The latitude and longitude coordinates are -35.20107 and 149.0966118, respectively. By considering these factors, you can make predictions about the price of this property based on its location, the size of the bedrooms, and the overall appeal of the suburb."*

Fig. 3. Unsupervised Context Learning Example

## D. *Framework for house price prediction*

This section will describe the framework we used for the study. Fig.2 illustrates the architecture of how we trained the mode and fine-tuning the GPT-2 model for house price prediction. First, we use dataset-2, X^'=[〖X^'〗_num, 〖X^'〗_cate, 〖X^'〗_context] to transform the features into a final property explanatory description dataset X^''=[〖X^'〗_(context+cate+num)] described above in Fig.3 , then we encode the final description features with max length 384 to get fixed length tokens that represents the final property explanatory token. Then we split the dataset X'' 90% of data for training and 10% of data for testing and apply them into GPT2 pre-trained model to continue training the model in an unsupervised manner. As a result, after the training processing, our framework generates the new model M'. According to the article [7], multi-head attention is a key component of the Transformer architecture that was introduced in the paper, because the multi-head attention mechanism allows the model to jointly attend to different parts of the input sequence, by computing multiple attention weights and feature representations in parallel. In other words, the multi-head attention can allow the language model to capture more complex relationships between the input tokens. In this experiment, we hope that the multi-head attention can help our model to leverage the importance of different aspects of the input property data and capture a different range of relationships between the features, such as the relationship between bedrooms, suburb, and prices.

The following is a brief description of multi-head self-attention learning we used for our experiment. A sequence of input matrix $X = [x_1,\ x_2, \dots x_n]$ represents $n$ features of one property. To perform multi-head attention, we use the input matrix $X$ to multiple heads using learned weight matrices $W_i$ and $i$ is the number of heads, and it can be formulated as following:

$$A_i = XW_i \tag{2}$$

where $A_i$ is the multi-head attention with the same dimension as $X$, but with a different learned weight matrix $W_i$ for each head. Next, for each head, the transformer model computes the attention weights and the corresponding weighted sum of the input sequence using the query *(*$Q^i$), key ($K^i$*),* and value ($V^i$) matrix respectively. Three different learned weight matrices $W_q^i$, $W_k^i$, and $W_v^i$ are used to compute the query, key, and value matrix with the input sequence $X$. They can be formulated as following,

$$Q^i = A_i W_q^i \tag{3}$$

$$K^i = A_i W_k^i \tag{4}$$

$$V^i = A_i W_v^i \tag{5}$$

After that, the transformer model has to get attention scores for each head $H_i$. The attention scores are computed by taking the dot product between the $Q^i$ and $K^i$, and scaling the result by the square root of the dimension of the $K^i$ matrix $d_k^i$. This is done to prevent the dot product from becoming too large or too small, which can cause problems during training. The attention scores can be formulated as following,

$$S^i = \frac{Q^i {K^i}^T}{\sqrt{d_k^i}} \tag{6}$$

The attention scores $S^i$ are then used to compute a set of attention weights for each head $H_i$, by applying a softmax function. And then the attention weights are then used to compute a weighted sum of the value vectors for each head $H_i$, to obtain the final representation for each token in the sequence $X$. It can be formulated as following,

$$\hat{A}_i(Q, K, V) = softmax(S^i)V \tag{7}$$

In GPT-2, the multi-head attention mechanism has two important parameters, one is the attention weights and the other is hidden states which play a critical role in computing the output representation for each token in the input sequence. Attention weights are a set of matrices that represent the importance of each input token for predicting the output at a given layer while hidden states can represent the understanding of input by a neural network at a given layer. For this experiment, we compare using head attentions and the last hidden states combining with basic features for house price prediction, the combining feature process can be defined as following (For equation (8), $i$ is the number of heads, for this experiment, $i$ is assigned a value of12. $A'_{last}$ represents last layer 12 head attentions combined with $X_{num}$ and $X_{cate}$. For equation (9), $S'_{last}$ represents final hidden state of each token of input sequences and $d$ is the total token of an input sequence).

$$A'_{last} = LayerNorm(Mean(H_i(a_1, \dots, a_i)) + X_{num} + X_{cate}) \tag{8}$$

$$S'_{last} = LayerNorm(Mean(S_i(s_1, \dots, s_d)) + X_{num} + X_{cate}) \tag{9}$$

### *E. Data Enhancement*

GPT-2 is a powerful generative language model that can be used for a variety of tasks, including data enhancement by providing extra information to the existing dataset. For our experiment, we use GPT-2 to continue training the model in an unsupervised in-context learning manner on the existing dataset $X'$. Once the model is trained, we generate new data based on a given seed text from dataset $X$. At the last, we use the generated data to enhance the existing dataset $X$ for the model fine-tuning training. During continuing unsupervised training, we typically use a pre-trained GPT-2 model as a starting point and fine-tune it for generating extra information based on six price range categories. An example of extra information as following for price range $750,000 ~ $900,000 (Fig. 4).

*"A comfortable residential property suitable for families, offering standard community services and private infrastructure in the suburb of Bruce."*

Fig. 4. Extra information for sample data enhancement

## IV. PERFORMANCE ANALYSIS

For analysing the performance of proposed approach, the fine tuning of GTP-2 model was done, leading to outputing the hidden states' parameters and attention weights from the model. In the context of transformer model, the attention weights are a set of vectors that represent the importance of each token of input context. During the training process, the model applies multi-heads to focus on different parts of the input sequence to get attentions weights for each head. In other words, attention weights can be used to determine which part of input tokens the model should concentrate on to make a more accurate prediction. In contrast, the hidden state is a vector that represents the model's understanding of the input context. For this experiment, we exam both parameters by using attentions weights from the default 12 multi-heads on the last layer as well as last layer states for analysing our proposed model.

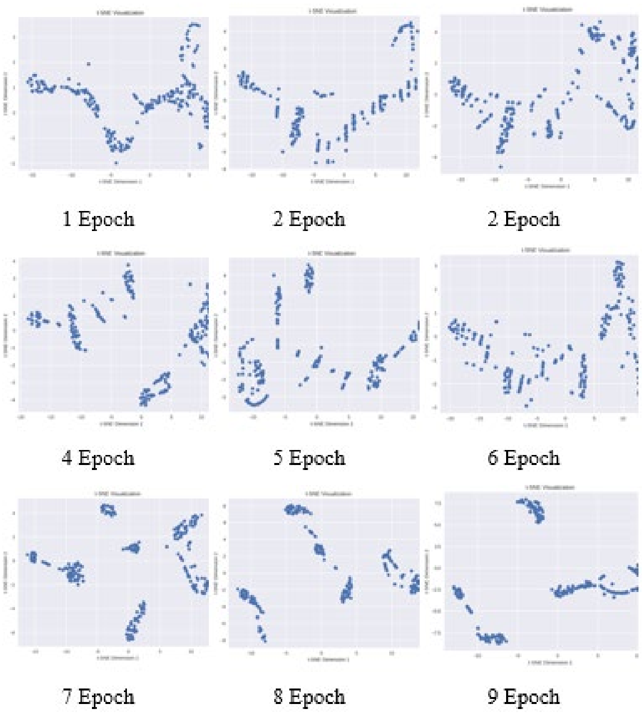


Fig. 5. Attention Visualisation

To get an understanding on how the attention weights changed during fine tuning process, we inspected the attention weights obtained from the model during training. As attention matrices are too large to visualize entirely, we use t-SNE visualisation to reduce the dimension of the attention matrix to 2 dimensions for visualisation. Each point represents a token in the attention matrix projected onto two dimensions, which are compressed using t-SNE algorithm. We found that this approach is helpful to analyse large attention weights. The points that are closer together indicate a higher level of similarity in the attention patterns learned by the model during the fine-tuning process. Fig. 5 shows a general visualization of attention weights for head-12 during fine-turning. As you can see that during first couple epochs, the attentions weights are scattered discretely, and then as the training progresses, they gradually converge into distinct clusters in different regions. In other words, the Fig. 5 provides insights of attention weights changed, where closer points indicate shared local relationships in terms of the original attention weights. Additionally, clusters of points indicate the number of distinct parts that the attention weights attend to within the input text. During epochs 7 to 9 fine-tuning, the attention weights appear to concentrate on about 4 to 5 distinct parts of the input text.

TABLE I. EXPERIMENTAL RESULTS

| *Model Name* | *Date Type* | *Epoch* | *MAE* | *MSE* |
|---|---|---|---|---|
| Decision Tree | $X_{num}$ | 1000 | 0.026 | 0.0056 |
| Random Forest | $X_{num}$ | 1000 | 0.023 | 0.0053 |
| MLP Regression | $X_{num}$ | 1000 | 0.033 | 0.0108 |
| K-Nearest Neighbors | $X_{num}$ | 1000 | 0.035 | 0.0065 |
| GPT-2 (hidden states) | Hidden states $X_{context}$, $X_{num}$ | 60 | 0.021 | 0.0015 |
| GPT-2 (attention weights) | Attention weights $X_{context}$ , $X_{num}$ | 60 | 0.022 | 0.0018 |

## V. CONCLUSIONS

In this study, we employed natural language processing techniques to develop a model for predicting real estate prices. Our approach involved extracting and comparing two parameters, namely hidden states and attention weights, from the last layer of the GPT-2 model. We combined these parameters with other features to enhance the accuracy of house price prediction. Furthermore, we enriched the original dataset X by incorporating additional information and explored the use of unsupervised in-context learning to further train the pre-trained GPT-2 model, aiming to achieve improved outcomes. The results of our experiment work demonstrated that our approach generated better results as compared to baseline machine learning models. Moreover, the findings indicated that both hidden states and attention weights can capture contextual understanding in real estate explanation text, but when the in-context input is short, hidden states appear to capture contextual understanding better. However, due to our limited resources and time constraints, we were unable to conduct in-depth analyses using larger language models such as GPT-3 or GPT-4. For future work, it would be worthwhile to investigate the use of in-context unsupervised learning in larger language models, evaluating whether such an approach can guide the language model towards more effective fine-tuning for real estate price prediction. Additionally, another exploration of research could involve utilising adapters instead of fine-tuning the entire language model, which would preserve the original language attention weights while only fine-tuning the adapter parameters. This adapter-based approach can be beneficial in designing sophisticated models for real estate price prediction.

## REFERENCES

[1] Z. Bragoudakis, M. Emiris, and M. Constantinescu, "Selected Review of the Empirical Literature on House Price Modelling and Forecasting: What Does the Literature Say?," SSRN Electronic Journal, 09/10 2016, doi: 10.2139/ssrn.3160945.

[2] A. D. Hjort, "House Price Prediction with Confidence: Empirical Results from the Norwegian Market," 2022.

[3] P.-Y. Wang, C.-T. Chen, J.-W. Su, T.-Y. Wang, and S.-H. Huang, "Deep Learning Model for House Price Prediction Using Heterogeneous Data Analysis Along With Joint Self-Attention Mechanism," IEEE access, vol. 9, pp. 55244-55259, 2021, doi: 10.1109/ACCESS.2021.3071306.

[4] J. Zaki, A. Nayyar, S. Dalal, and Z. H. Ali, "House price prediction using hedonic pricing model and machine learning techniques," Concurrency and computation, vol. 34, no. 27, p. n/a, 2022, doi: 10.1002/cpe.7342.

[5] S. Ozogur Akyuz, B. Eygi Erdogan, O. Y?ld?z, and P. Karaday? Atas, "A Novel Hybrid House Price Prediction Model," (in eng), Computational economics ,, 2022.

[6] A. S. Temur, M. Akgun, and G. Temur, "PREDICTING HOUSING SALES IN TURKEY USING ARIMA, LSTM AND HYBRID MODELS," Journal of business economics and management, vol. 20, no. 5, pp. 920-938, 2019, doi: 10.3846/jbem.2019.10190.

[7] A. Vaswani et al., "Attention Is All You Need," 2017.

[8] A. a. N. Radford, Karthik and Salimans, Tim and Sutskever, Ilya, "Improving language understanding by generative pre-training," 2018.

[9] J. Pfeiffer et al., "AdapterHub: A Framework for Adapting Transformers," 2020, doi: 10.48550/arxiv.2007.07779.

[10] F. Petroni et al., "Language Models as Knowledge Bases?," 2019, doi: 10.48550/arxiv.1909.01066.

[11] A. K. Lampinen et al., "Can language models learn from explanations in context?," 2022, doi: 10.48550/arxiv.2204.02329.

[12] J. Coda-Forno, M. Binz, Z. Akata, M. Botvinick, J. X. Wang, and E. Schulz, "Meta-in-context learning in large language models," 2023, doi: 10.48550/arxiv.2305.12907.

[13] S. Vashishth, S. Upadhyay, G. S. Tomar, and M. Faruqui, "Attention Interpretability Across NLP Tasks," 2019, doi: 10.48550/arxiv.1909.11218.

[14] J. Liu, D. Shen, Y. Zhang, B. Dolan, L. Carin, and W. Chen, "What Makes Good In-Context Examples for GPT-3?," 01, 2022.

[15] A. Trott, C. Xiong, and R. Socher, "Interpretable Counting for Visual Question Answering," 2017, doi: 10.48550/arxiv.1712.08697..

[16] S. Jain and B. C. Wallace, "Attention is not Explanation," 2019, doi: 10.48550/arxiv.1902.10186.

[17] Zhao, Y., Chetty, G., & Tran, D. (2019). Deep Learning with XGBoost for Real Estate Appraisal. In 2019 IEEE Symposium Series on Computational Intelligence, SSCI 2019 (pp. 1396-1401). [9002790] (2019 IEEE Symposium Series on Computational Intelligence, SSCI 2019). IEEE, Institute of Electrical and ElectronicsEngineers. https://doi.org/10.1109/SSCI44817.2019.9002790

[18] Y. Li, P. Branco, and H. Zhang, "Imbalanced Multimodal Attention-Based System for Multiclass House Price Prediction," Mathematics (Basel), vol. 11, no. 1, p. 113, 2023, doi: 10.3390/math11010113.

[19] A. Radford, J. Wu, R. Child, D. Luan, D. Amodei, and I. Sutskever, "Language Models are Unsupervised Multitask Learners," ed: OpenAI, 2019.

[20] Y. Zhao, G. Chetty and D. Tran, "Deep Learning for Real [2Estate Trading," in 2022 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE), Gold Coast, Australia, 2022 pp. 1-7. doi: 10.1109/CSDE56538.2022.10089222.