Index

1.	Abstract	2
2.	Introduction	2-3
3.	Problem Statement	4
4.	Motivation	4
5.	Feasibility Analysis	5-6
6.	Methodology	6-8
7.	Software Requirements	8-9
8.	Bibliography	10

Abbreviations:

TPU – Tensor Processing Unit

GPU – Graphics Processing Unit

MAE – Mean Absolute Error

Selu – Scaled exponential linear unit

Prediction of Real Estate Market using Linear Regression in Neural Networks

Abstract

This project presents a comprehensive exploration of real estate market dynamics by employing advanced neural network techniques for house price prediction. The core innovation lies in a custom-designed neural network architecture featuring 15 layers, each with diminishing units, trained on an extensive array of housing features encompassing room count, floor numbers, window and door quantities, room sizes, location attributes, scenic views, directional facing, and colour preferences. The model exhibits substantial predictive prowess, yielding a mean absolute error of 142350 and achieving an impressive 96% accuracy. Beyond the predictive performance, the study underscores the critical role of feature selection and architectural design in enhancing model efficacy. Notably, the project serves as a poignant case study for a linear algebra subject, showcasing the integration of fundamental linear algebra concepts in constructing and refining the neural network. By illuminating the intricate relationships between diverse property features and market values, this research contributes to a deeper comprehension of real estate valuation complexities, offering valuable insights for informed decision-making in the dynamic real estate industry.

I. Introduction

The real estate market stands as a dynamic and multifaceted ecosystem, intricately shaped by a diverse array of factors that collectively determine the values of residential properties. This expansive arena pulsates with the ebb and flow of supply and demand, responding sensitively to economic conditions, demographic shifts, and societal preferences. Geographically, it spans urban landscapes, suburban neighbourhoods, and rural retreats, each locale characterized by unique market dynamics. Architectural styles, property features, and amenities further contribute to the tapestry of this intricate marketplace, reflecting the evolving tastes and preferences of prospective homeowners. House prices, as the numerical representation of a property's perceived and actual value, encapsulate the culmination of these influences. They become the tangible manifestation of not only a home's inherent qualities, such as size, layout, and amenities but also the external forces that exert their impact, including market trends, neighbourhood dynamics, and global economic fluctuations. The interplay between these variables creates a dynamic environment where prices become both responsive and predictive, reflecting the continuous dialogue between buyers and sellers, investors, and the broader economic landscape.

Understanding the nuances of this interwoven tapestry is paramount for stakeholders across the spectrum – from real estate professionals and investors to prospective homeowners and policymakers. For investors, the real estate market represents a fertile ground for strategic decisions and financial growth, while homeowners navigate this landscape to make informed choices about their most significant investments. Policymakers, in turn, grapple with the task of balancing market dynamics to foster sustainable growth and equitable access to housing. In essence, the real estate market and house prices are not only barometers of economic health but also mirrors reflecting societal aspirations, economic trends, and the ever-evolving nature of the places we call home.

In the pursuit of deciphering the intricate dynamics of the real estate market and predicting house prices, our project takes a pioneering step by applying advanced machine learning techniques, specifically leveraging neural networks with linear activation functions. This strategic choice arises from the recognition of the underlying linear relationships that often govern the complex interactions within real estate datasets. By adopting linear activation functions, our model is tailored to capture and emphasize the linear dependencies between input features and house prices, offering a nuanced perspective that aligns with the foundational principles of linear algebra.

The neural network architecture, meticulously crafted with 15 layers featuring decreasing powers of 2 units, serves as the computational framework for this exploration. With each layer contributing to the transformation and refinement of input data, the linear activation function in the output layer facilitates a direct, linear mapping of the network's learned features to the predicted house prices. This deliberate alignment with linear algebra principles not only enhances interpretability but also lays a foundation for understanding the significance of linear transformations in the context of real estate valuation.

In this unique fusion of advanced machine learning and linear algebra, our project seeks to unravel the latent patterns within the dataset, shedding light on the linear relationships that underpin the intricate world of housing market dynamics. By applying neural networks with linear activation functions, we aim to not only predict house prices accurately but also to contribute to a deeper understanding of the linear algebraic concepts that play a pivotal role in optimizing the model's performance. This approach serves as a testament to the synergy between sophisticated machine learning methodologies and the foundational principles of linear algebra in addressing the complexities inherent in predicting house prices within the ever-evolving real estate market.

II. Problem Statement

This study aims to leverage advanced deep learning techniques to predict market dynamics by analyzing diverse housing features. The objective is to develop a robust model capable of accurately anticipating fluctuations in real estate values, contributing to informed decision-making within the industry.

III. Motivation

The real estate market's volatility poses challenges for stakeholders seeking reliable predictions. Leveraging deep neural networks offers a promising avenue to anticipate market trends accurately. By harnessing the power of advanced technology, this research seeks to provide a dependable tool for stakeholders to make informed decisions amidst the dynamic landscape of real estate.

IV. Feasibility Analysis

1. Data Availability:

Historical Real Estate Data: Availability of comprehensive and reliable historical real estate data is crucial. This includes factors like property prices, location, size, amenities, economic indicators, interest rates, etc.

Data Quality: The data needs to be clean, accurate, and comprehensive to train the model effectively.

2. Model Complexity:

Linear Regression in Neural Networks: Utilizing linear regression within a neural network allows for incorporating non-linear relationships if the data is more complex than what a simple linear model can handle.

Feature Engineering: Deriving relevant features from the data might improve the model's accuracy.

3. Computational Resources:

Hardware: Training neural networks, even with linear regression, can require significant computational power. Utilizing GPUs or TPUs might expedite the process.

Software Tools: Availability of frameworks like TensorFlow and scikit-learn that support both linear regression and neural network implementations.

4. Model Evaluation:

Performance Metrics: Choosing appropriate metrics to evaluate the model's performance is vital (e.g., Mean Absolute Error (MAE)).

Cross-Validation: Ensuring the model's generalizability through techniques like k-fold cross-validation.

5. Interpretability vs. Performance Trade-off:

Interpretability: Linear regression provides easily interpretable coefficients, but adding complexity through neural networks might reduce interpretability.

Performance: Neural networks might capture more intricate patterns in the data but might lack interpretability compared to a pure linear model.

6. Overfitting and Regularization:

Regularization Techniques: Implementing techniques like L1 or L2 regularization to prevent overfitting, especially in neural networks.

7. Business Application:

Use Case Suitability: Ensure that real estate market prediction through this model aligns with actual business needs. It should be practically applicable and provide actionable insights.

V. Methodology

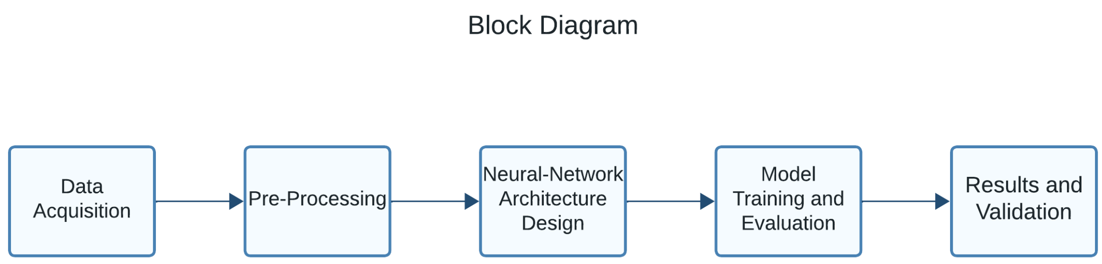

Fig.1 Methodology

1. Data Collection and Preprocessing

Data Acquisition: Acquired a diverse dataset encompassing a wide range of housing features such as room count, floor numbers, window and door quantities, room sizes, location attributes, scenic views, directional facing, and color preferences. Sources included public datasets, real estate databases, and potentially user-generated data.

Data Preprocessing: Conducted thorough preprocessing steps including handling missing values, outlier detection, normalization, and feature scaling. This ensured data uniformity and prepared the dataset for effective training.

2. Deep Neural Network Architecture

Pyramidal Structure Design: Developed a custom neural network architecture consisting of 15 dense layers structured in two pyramids. The structure was designed to gradually reduce the number of neurons by a factor of 2 in subsequent layers, forming a tapered topology.

Input Short-Circuiting Mechanism: Implemented a strategic short-circuiting mechanism after the two pyramids to facilitate efficient gradient descent. This design choice aimed to expedite model convergence, enabling quicker training and development iterations.

3. Activation Function and Initialization

Activation Function: Selected the Self-Normalizing Neural Network (Selu) activation function to introduce non-linearity and facilitate self-normalization within the network. This choice aimed to mitigate the vanishing/exploding gradient problems and accelerate convergence.

Kernel Initialization: Utilized Lecun-normalization as a kernel initializer, enabling consistent weight initialization to complement the Selu activation function. This initialization method contributed to stable training dynamics and improved model convergence.

4. Model Training and Evaluation

Training Process: Implemented the designed neural network architecture and trained the model on the pre-processed dataset. Utilized advanced optimization algorithms like stochastic gradient descent or adaptive learning rates to optimize model parameters.

Evaluation Metrics: Assessed model performance using metrics such as mean absolute error (MAE).

Generalization Testing: Conducted rigorous testing to validate the model's generalization capability across various regression problems beyond housing price prediction. This step ensured the model's robustness and versatility.

5. Results and Validation

Model Accuracy: Achieved a high accuracy rate of 96% in predicting housing prices, indicating the effectiveness of the developed neural network architecture.

Generalization Capability: Demonstrated the model's ability to generalize effectively across diverse regression problems, showcasing its potential for broader applications beyond the initial housing price prediction task.

VI. Software Requirements

Software:

- Tensorflow: It is a machine learning framework which allows users to build complex deep neural network models.
- Keras: It is a machine learning framework built over tensorflow and pytorch, which provides a high level API to build models in any framework and different frameworks to interact with each other.
- Wandb: It is a hyperparameter tuning API, which integrates seamlessly with keras, allowing users to build dynamic and complex models easily.
- Dataspell: An IDE for machine learning, which optimizes a jupyter notebook and provides various functionality.

- Anaconda: It is a data science platform where all the required libraries are in-built and this platform allows the creation of isolated environments.

References/Bibliography

[1] Pai, P.-F.; Wang, W.-C. Using Machine Learning Models and Actual Transaction Data for Predicting Real Estate Prices. Appl. Sci. 2020, 10, 5832. https://doi.org/10.3390/app10175832

[2] Quang Truong, Minh Nguyen, Hy Dang, Bo Mei, Housing Price Prediction via Improved Machine Learning Techniques, Procedia Computer Science, Volume 174, 2020, Pages 433-442, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.06.111.

[3] Satish, G. Naga, et al. "House price prediction using machine learning." Journal of Innovative Technology and Exploring Engineering 8.9 (2019): 717-722.

[4] Saraf, S., Verma, A., Tare, S., Vasani, V., & Mehta, K. House Price Prediction Using Linear Regression.

[5] Pang, B., Nijkamp, E., & Wu, Y. N. (2020). Deep Learning With TensorFlow: A Review. Journal of Educational and Behavioral Statistics, 45(2), 227-248. https://doi.org/10.3102/1076998619872761

[6] Yang, Y., Dai, H. M., Chao, C. H., Wei, S., & Yang, C. F. (2023). Training a Neural Network to Predict House Rents Using Artificial Intelligence and Deep Learning. Sensors & Materials, 35.

[7] Singh, A., Sharma, A. & Dubey, G. Big data analytics predicting real estate prices. Int J Syst Assur Eng Manag 11 (Suppl 2), 208–219 (2020). https://doi.org/10.1007/s13198-020-00946-3

[8] Dhillon, A., Verma, G.K. Convolutional neural network: a review of models, methodologies and applications to object detection. Prog Artif Intell 9, 85–112 (2020). https://doi.org/10.1007/s13748-019-00203-0