

# **Final Project**

---

## **Team Members:**

**1. Abir Saha**

**Semo Id : S02071626**

**2. Xiao pu**

**Semo Id : S01970474**

**3. Rohith Singh**

**Semo Id : S02062249**

---

**For project 3, from 2 choices, we choose the kaggle competition. Here is our full report.**

## **House prediction using ANN Model:**

**Competition link :** <https://www.kaggle.com/competitions/house-prices-advanced-regression-techniques/overview>

**Notebook link :** <https://www.kaggle.com/code/abirsaha123456/ann-model>

## **1. Objective**

The objective of this project was to predict home prices while training a deep learning model using the House Prices: Advanced Regression Techniques dataset from Kaggle. As we know that regression problem can be solved by the ANN algorithm. So in this project we try to solve it by ANN and kaggle competition accept this one.

## **2. Data Preprocessing**

To prepare the data for training, we followed these steps:

- Dropped the Id column. (Unnecessary column get idea from others)
- Missing values were imputed by taking the mean for numeric columns and the mode for categorical ones.
- One-hot encoding was used for categorical variables.

- Features were standardized using StandardScaler to improve neural network performance

### **3. ANN Architecture**

The model consists of:

- Input Layer: 128 neurons, ReLU activation
- Hidden Layer 1: 64 neurons, ReLU activation
- Hidden Layer 2: 32 neurons, ReLU activation
- Output Layer: 1 neuron, linear activation (for regression)

Compiled with Adam optimizer, MSE loss function, and MAE as the evaluation metric.

### **4. Training Setup**

- The data was divided into 70% training and 30% validation data.
- The model trained for 100 epochs using a batch size of 32.
- The validation data was used to monitor generalization or performance on previously untrained data.

### **5. Evaluation Results**

The final evaluation on the validation set:

- Training MAE: 10,334
- Validation MAE: 23,709
- RMSE: ~37,842
- R<sup>2</sup> Score: ~0.87

The model shows good performance and generalization without major overfitting.

### **6. Visualization of Learning Curves**

Two plots were generated to monitor model training progress over 100 epochs:

- Training and Validation Loss: This shows how well the model minimizes the error. Both curves drop steeply in the first 10 epochs and then stabilize, indicating effective learning.
- Training and Validation MAE: Shows the mean absolute error over time. The validation MAE stabilizes after about 60 epochs, suggesting convergence.

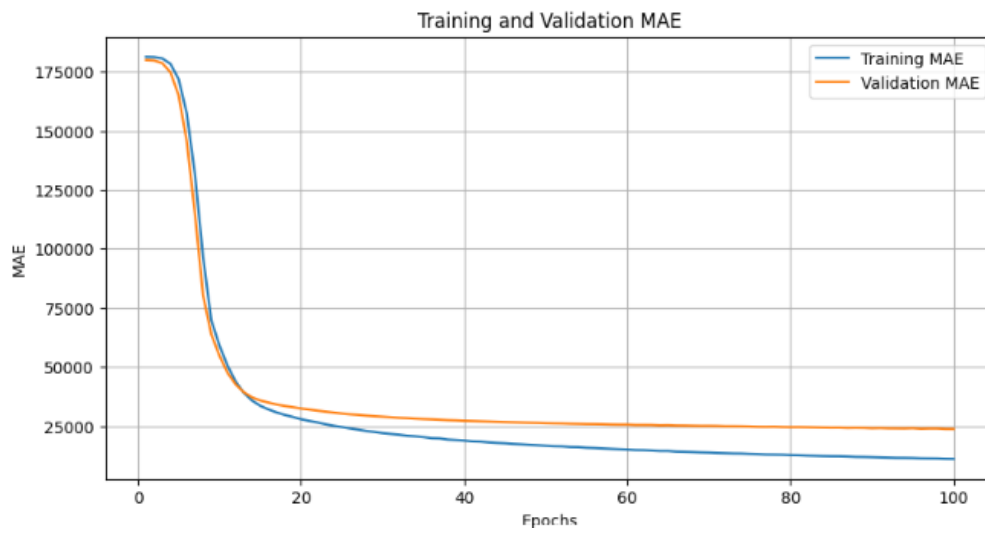
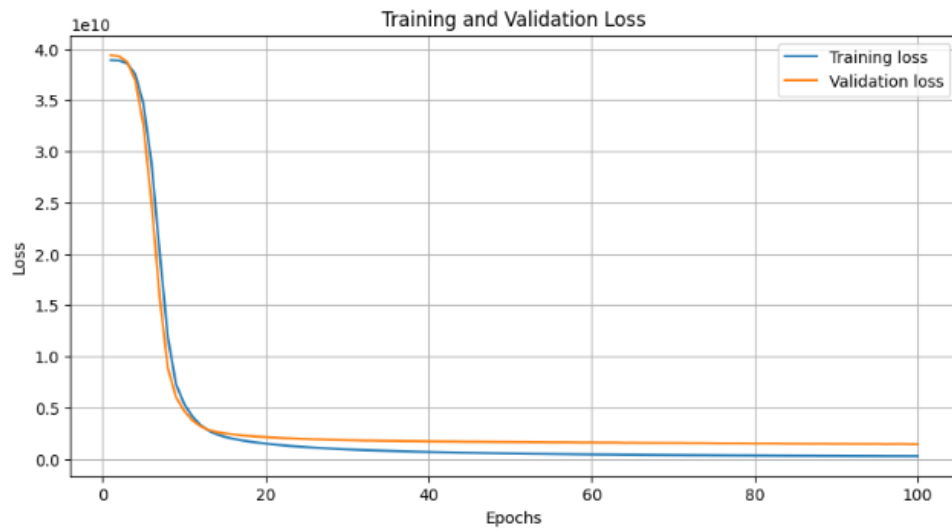
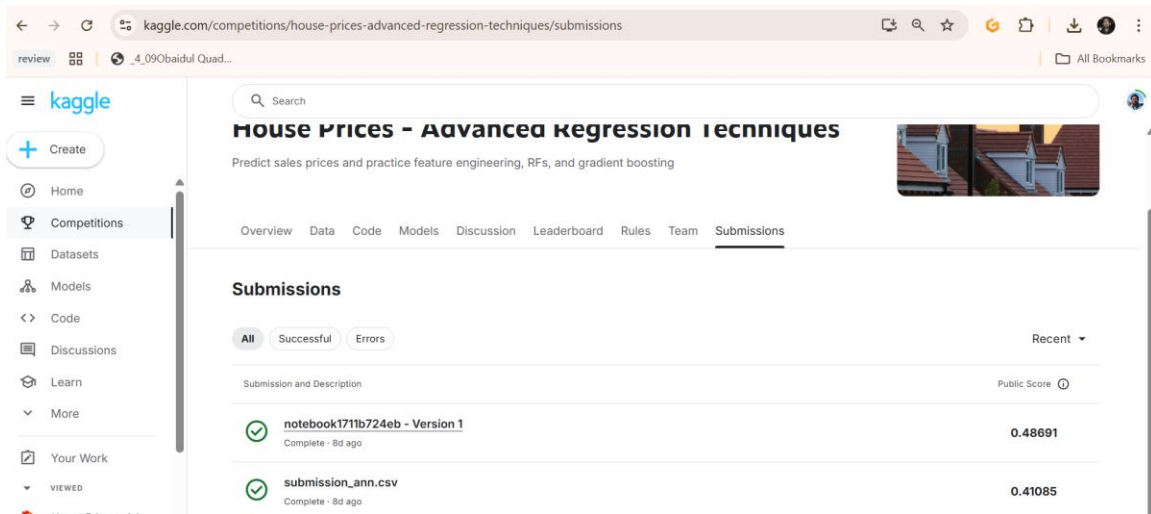


Figure: Training and Validation Loss and MAE over 100 Epochs



## 7. Prediction and Submission

The test dataset underwent the same processing steps, predictions were made, and 'submission.csv' containing the 'Id' and 'SalePrice' columns for submission to Kaggle was generated and saved.

After submitting the csv file the public score was 0.41085 and the leaderboard position is 4351.



The screenshot shows the Kaggle competition page for 'House Prices - Advanced Regression Techniques'. The page is titled 'House Prices - Advanced Regression Techniques' with a subtitle 'Predict sales prices and practice feature engineering, RFs, and gradient boosting'. The 'Submissions' tab is selected, showing a list of submissions. The table has columns for 'Submission and Description' and 'Public Score'. Two submissions are visible: 'notebook1711b724eb - Version 1' with a score of 0.48691 and 'submission\_ann.csv' with a score of 0.41085. Both submissions are marked as 'Complete' and '8d ago'.

Submission and Description	Public Score
 <b>notebook1711b724eb - Version 1</b> Complete · 8d ago	0.48691
 <b>submission_ann.csv</b> Complete · 8d ago	0.41085

The screenshot shows the Kaggle website interface. On the left is a sidebar with navigation links: Home, Competitions, Datasets, Models, Code, Discussions, Learn, More, Your Work, and a 'VIEWED' section with recent activity. The main content area is titled 'Leaderboard' and includes a search bar with 'abir' entered. Below the search bar, a message states: 'This leaderboard is calculated with all of the test data.' A table displays the leaderboard data:

#	Team	Members	Score	Entries	Last	Join
4351	abir #3		0.41085	3	8d	

Below the table, a message reads: 'Your Best Entry! Your submission scored 0.48691, which is not an improvement of your previous score. Keep trying!' The top navigation bar includes links for Overview, Data, Code, Models, Discussion, Leaderboard, Rules, Team, and Submissions. The top right corner shows a search bar and a user profile icon.

## 8. Conclusion

This project shows that a model leveraging deep learning, like ANN, can be used in structured regression problems.