Final Project

Team Members:

1. Abir Saha	**2. Xiao pu**	**3. Rohith Singh**
Semo Id : S02071626	**Semo Id : S01970474**	**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^2 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.

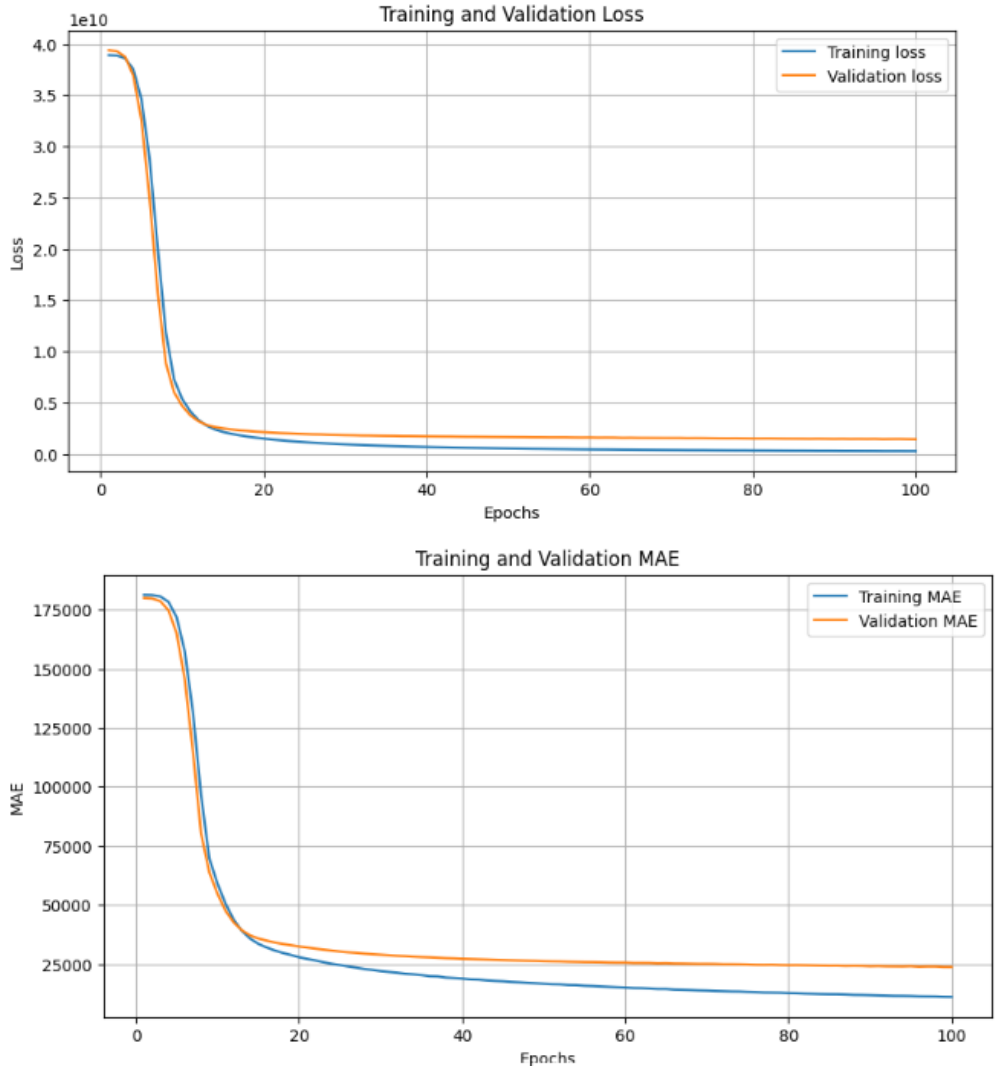
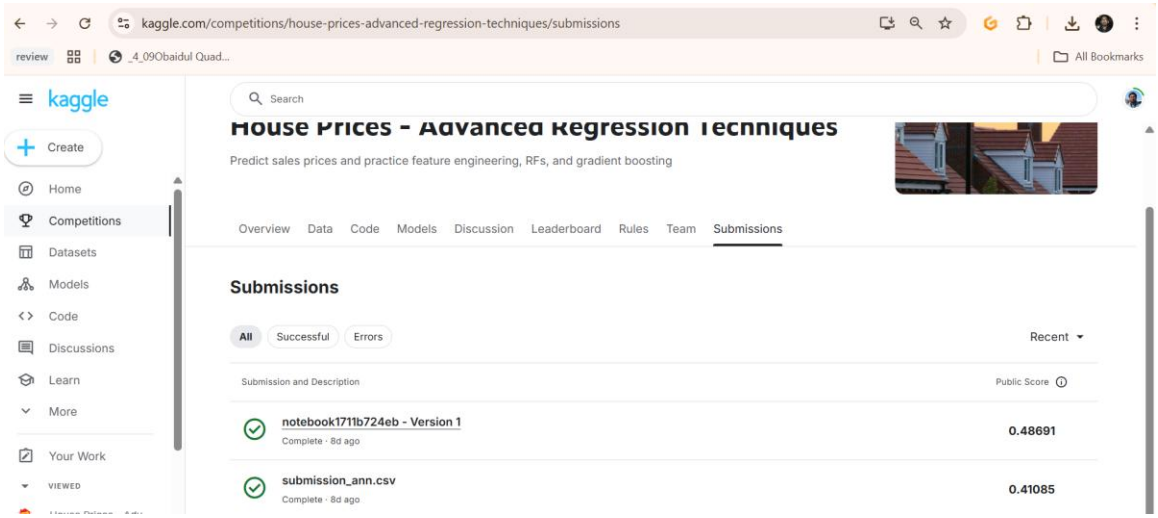


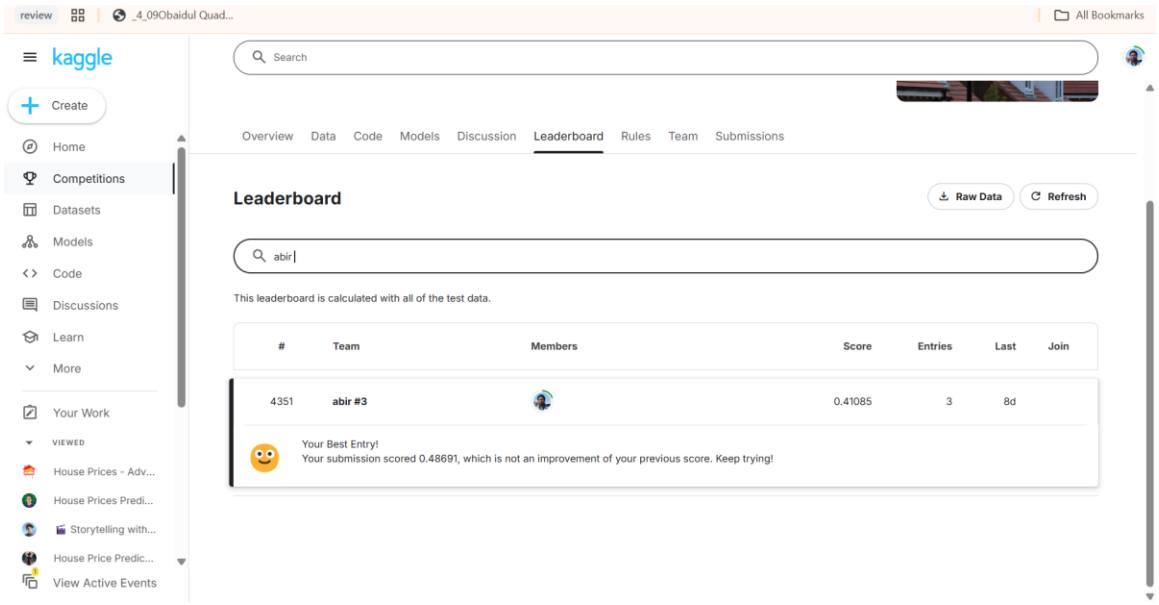
Figure: Training and Validation Loss and MAE over 100 Epochs

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.





8. Conclusion

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