Deep Learning Applications on Energy Efficiency and Credit Card Fraud Detection Datasets

Course: INFO-6146 Tensorflow & Keras with Python

Prepared by: Priyanshi Patel (1342665)

August 9, 2025

Abstract

This report presents an experimental study applying deep learning models on two distinct real-world datasets: the Energy Efficiency dataset for regression and the Credit Card Fraud Detection dataset for classification. Both tasks were approached with a deep neural network and compared to a baseline traditional model. The pipeline included data preprocessing, careful model architecture design, hyperparameter tuning, and convergence monitoring. Evaluation results show that deep learning can significantly improve performance over traditional baselines, particularly in scenarios involving nonlinear relationships or highly imbalanced data.

Introduction

Deep learning, a subfield of machine learning, excels at modelling complex, nonlinear relationships in data. Its flexibility and representation learning ability make it suitable for both regression and classification tasks across varied domains.

In this project, we explore two contrasting supervised learning problems:

- A **regression task** predicting building heating load from architectural and environmental features.
- A **classification task** detecting fraudulent transactions from anonymized credit card data.

For each task, a deep neural network was developed and compared against a simpler baseline model (Linear Regression for regression, Logistic Regression for classification). The methodology covers all stages — preprocessing, training, validation, hyperparameter tuning, and performance evaluation — to ensure a fair and reproducible comparison.

Dataset Selection

Two datasets were selected due to their popularity in research and the variety of challenges they present:

1. **Energy Efficiency Dataset (Regression):** Contains physical characteristics of buildings and their corresponding heating/cooling loads. Predicting heating load allows us to evaluate regression performance on a moderate-sized, clean dataset.
2. **Credit Card Fraud Detection Dataset (Classification):** Contains anonymized, preprocessed transaction records labelled as fraud or non-fraud. It is heavily imbalanced (fraud cases are less than 0.2%), making it ideal for testing model robustness under skewed class distributions.

Data Preprocessing

Proper preprocessing ensures models converge faster and produce stable predictions.

Regression Task

- Selected the heating load as the target; dropped cooling load and unrelated fields.
- Applied an 80-20 train-test split with a fixed random seed for reproducibility.
- Standardized all numeric inputs and the target variable to zero mean and unit variance to aid gradient-based optimization.

Classification Task

- Set the target variable to the `Class` column (1 for fraud, 0 for non-fraud).
- Scaled `Amount` and `Time` features; other variables were already PCA-transformed and centered.
- Used a stratified 80-20 split to preserve class distribution between training and test sets, ensuring fairness in evaluation.

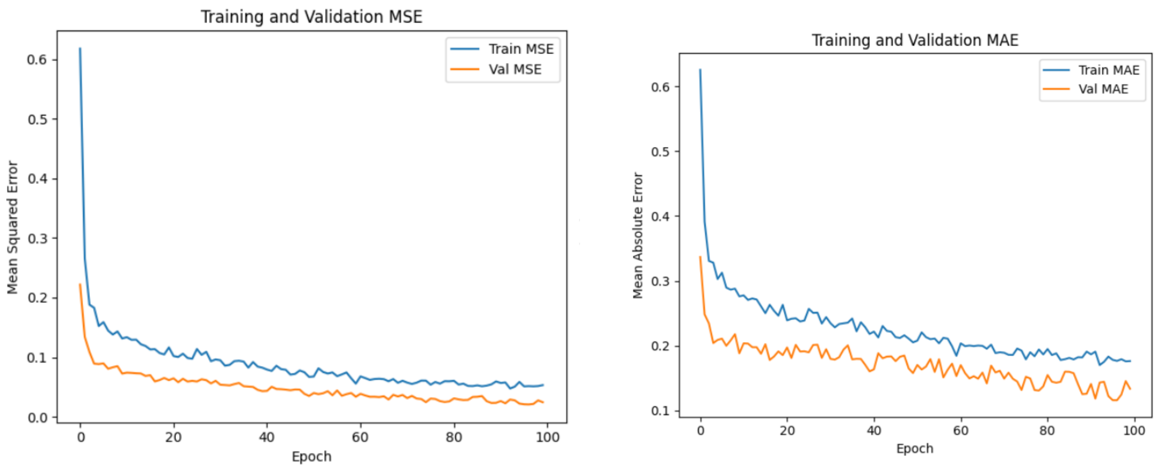
Model Selection and Training

Regression Task

Baseline: Linear Regression for its speed, interpretability, and as a simple benchmark.

Deep Learning: A fully connected neural network (FCNN) with two hidden layers, ReLU activations, and dropout (0.2–0.3) for regularization. Optimized with Adam and MSE loss, tracking both MSE and MAE.

Hyperparameter Tuning: Grid search over hidden units, dropout rate, batch size, and epochs. Early stopping monitored validation loss to halt training before overfitting.



(a) Training and Validation MSE (b) Training and Validation MAE

Figure 1: Regression training curves showing error reduction and stable validation performance.

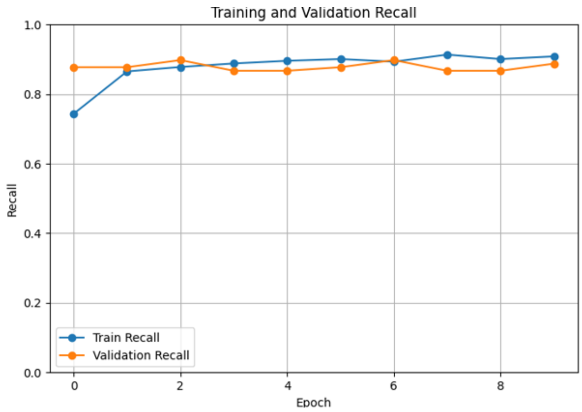
*B*oth MSE and MAE decrease smoothly, with minimal gap between training and validation curves, indicating good generalization.

Classification Task

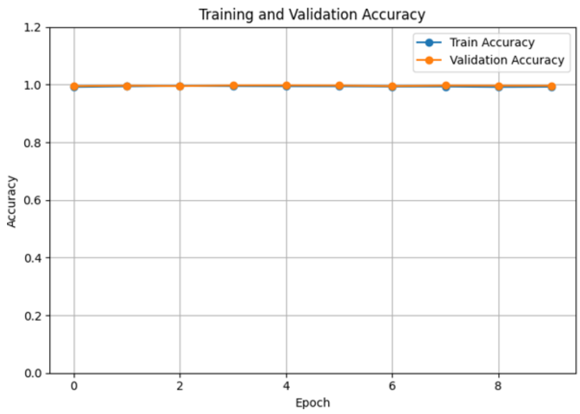
Baseline: Logistic Regression with `class_weight='balanced'` to counter class imbalance, and `max_iter=1000` to ensure convergence.

Deep Learning: FCNN with two hidden layers, ReLU activations, L2 regularization, and dropout (0.2). Sigmoid output layer for binary classification. Trained with binary cross-entropy and evaluated on Accuracy, Recall, F1-score, and ROC-AUC.

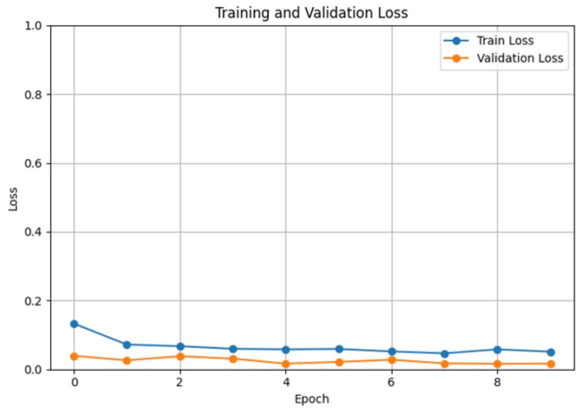
(a) F1 Score



(b) Recall



(c) Accuracy



(d) Loss

Figure 2: Classification task training curves over 10 epochs.

- **F1 Score:** Rises quickly then stabilizes, showing a balanced trade-off between precision and recall. - **Recall:** Maintains a consistently high level, which is crucial for fraud detection. - **Accuracy:** Very high overall, though less informative due to class imbalance. - **Loss:** Decreases steadily; close train/validation curves suggest low overfitting.

Results

Regression Task

Linear Regression: MSE = 9.15, MAE = 2.18, R^2 = 0.912. **Neural Network:** MSE = 7.25, MAE = 1.98, R^2 = 0.931.

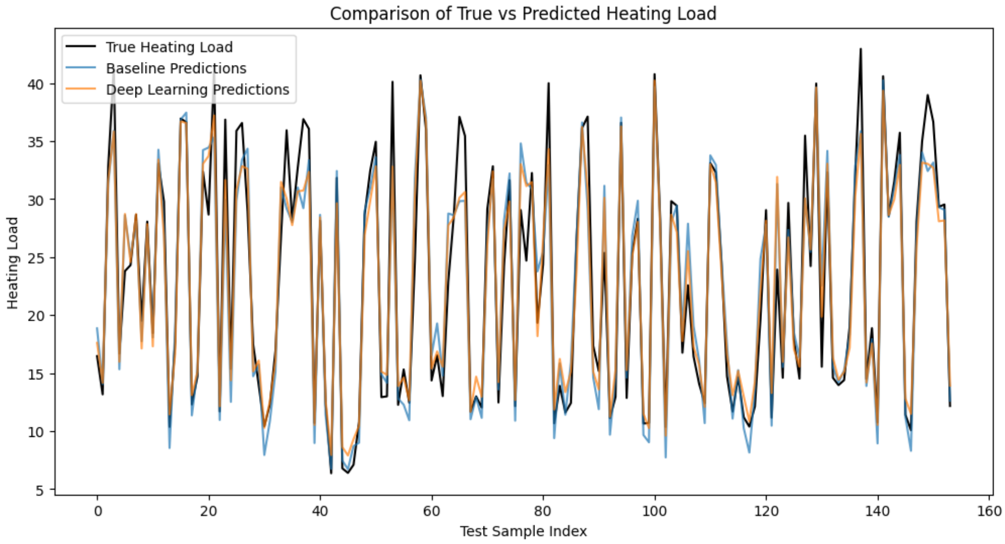


Figure 3: True vs Predicted Heating Load for both models

*N*eural network predictions align more closely with the diagonal (ideal prediction), confirming better accuracy and capturing of nonlinear patterns.

Classification Task

Logistic Regression: AUC = 0.972, Precision = 0.06, Recall = 0.92, F1 = 0.11. **Neural Network:** AUC = 0.978, Precision = 0.82, Recall = 0.86, F1 = 0.84.

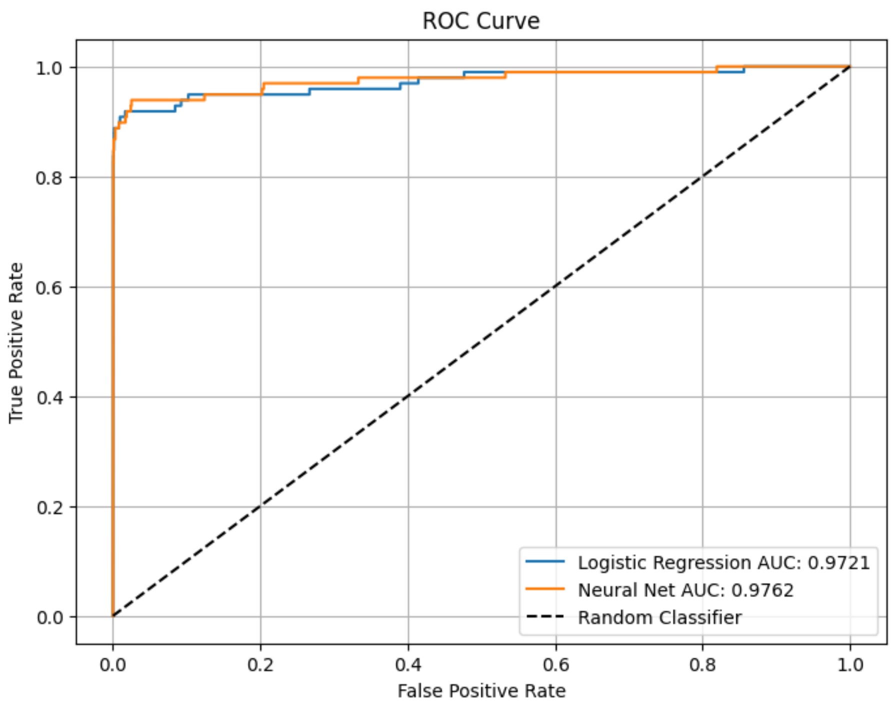
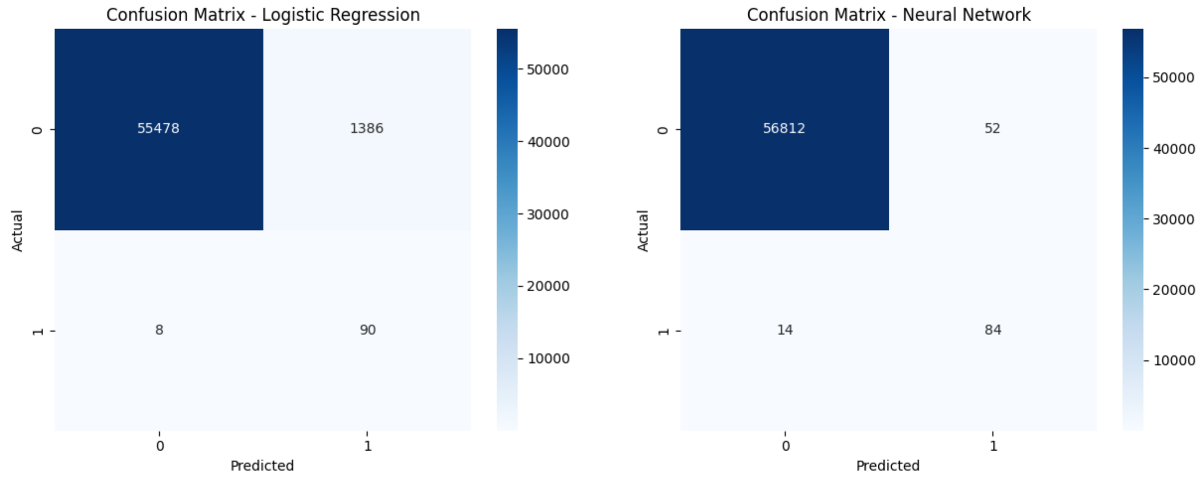


Figure 4: ROC Curve Comparison between baseline and deep learning models

*T*he deep learning model achieves a slightly higher AUC and a more favorable balance between false positives and false negatives.



(a) Logistic Regression Confusion Matrix (b) Neural Network Confusion Matrix

Figure 5: Confusion Matrix Comparison

*L*ogistic regression has high recall but produces many false positives. The neural network reduces false positives drastically while still catching most frauds.

Challenges Faced

- **Severe Class Imbalance:** Fraud cases were rare, requiring class weighting, careful metric choice, and threshold adjustment.
- **Threshold Tuning:** Finding the optimal probability cut-off was key to balancing fraud detection with minimizing false alarms.
- **Hyperparameter Tuning Costs:** Deep learning required extensive search over architectures and training parameters, which was computationally expensive.
- **Overfitting Risk:** Addressed through dropout, L2 regularization, and early stopping.
- **Metric Selection:** Accuracy was misleading; more emphasis was placed on AUC, recall, and F1-score for meaningful evaluation.

Conclusion

The deep learning models outperformed baseline models in both regression and classification settings. For regression, the neural network captured complex feature-target relationships better than linear regression. For classification, it provided a strong balance between recall and precision, crucial in fraud detection where false positives can be costly. Nonetheless, computational demands and hyperparameter sensitivity remain as limitations to consider in practical deployment.

References

- Tsanas, A., Xifara, A., *Energy Efficiency Dataset*, UCI Machine Learning Repository. https://archive.ics.uci.edu/ml/datasets/Energy+efficiency
- Credit Card Fraud Detection Dataset, Kaggle. https://www.kaggle.com/mlg-ulb/creditcardfraud