Data Science Salary Fairness Classification Project

Objective

The aim of this project is to classify whether a given salary is fair or underpaid based on features such as job title, location, experience, and other relevant factors. Using deep learning, we explore different neural network architectures to find an optimal model that achieves strong classification performance.

Dataset Description

The dataset consists of labeled examples indicating whether a salary is considered "fair" or "underpaid". The input features were preprocessed and transformed into a dense numeric format suitable for input into neural network models.

Model Variations

We experimented with three different model architectures:

Model V1 (Baseline Model)

- Dense Layer: 64 units, ReLU activation
- Dense Layer: 32 units, ReLU activation
- Output Layer: 1 unit, Sigmoid activation
- Optimizer: Adam (lr=0.001)
- Epochs: 25
- Batch size: 64

Model V2 (Deeper Network without Regularization)

- Dense Layer: 128 units, ReLU
- Dense Layer: 64 units, ReLU
- Dense Layer: 32 units, ReLU
- Output Layer: 1 unit, Sigmoid
- Same optimizer and training settings as Model V1

Model V3 (Regularized Model)

- Dense Layer: 64 units, ReLU, L2 regularization (0.01)
- Dropout: 0.5
- Dense Layer: 32 units, ReLU, L2 regularization (0.01)
- Dropout: 0.5
- Output Layer: 1 unit, Sigmoid

Evaluation Metrics

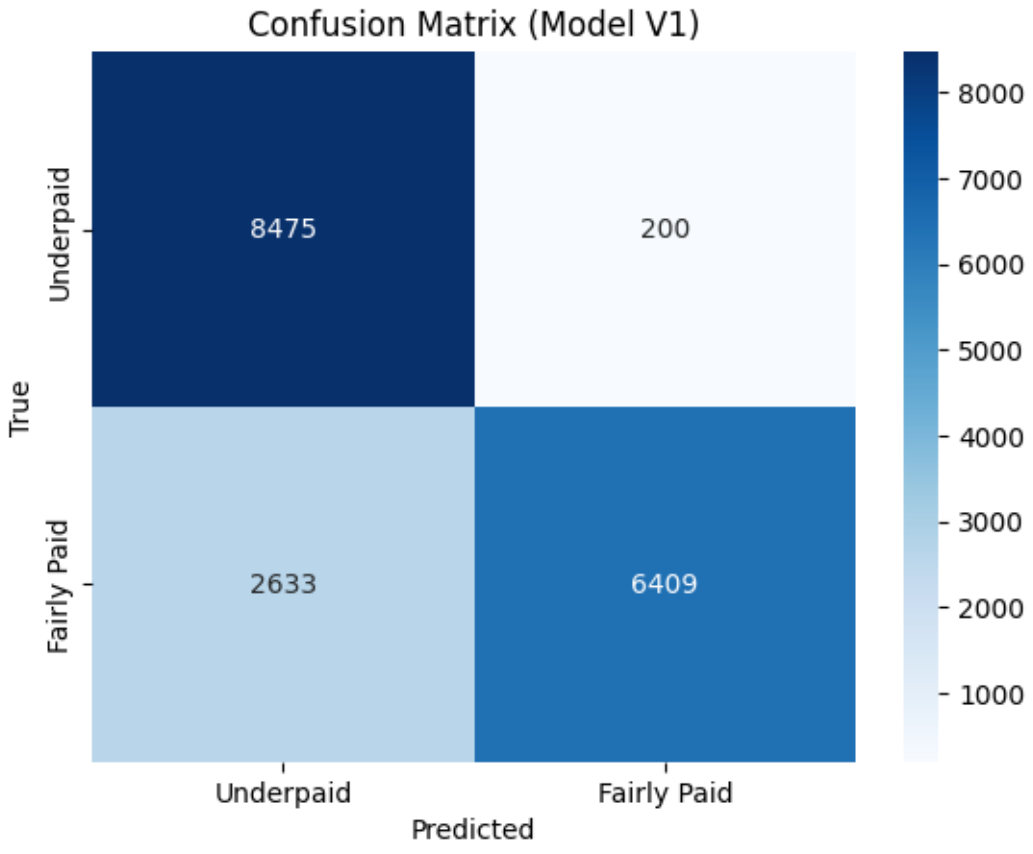
Each model was evaluated using:

- **Test Accuracy**
- **Loss Curves (Train vs Validation)**
- **Confusion Matrix**
- **ROC Curve & AUC**

Results

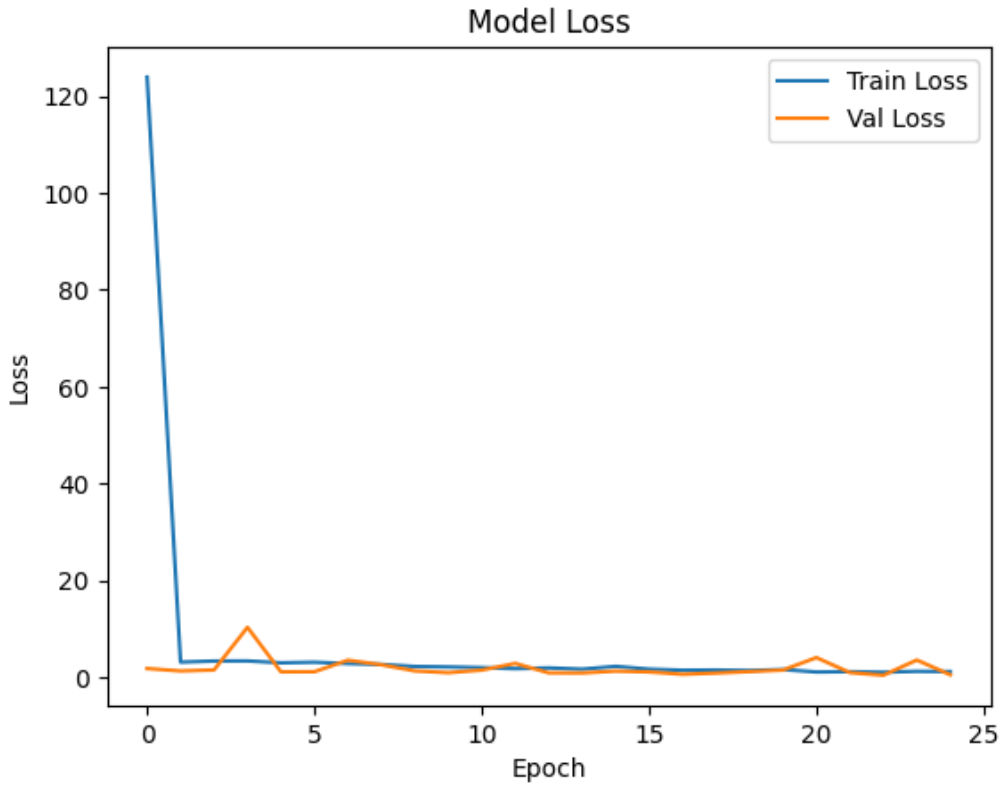
Model V1:

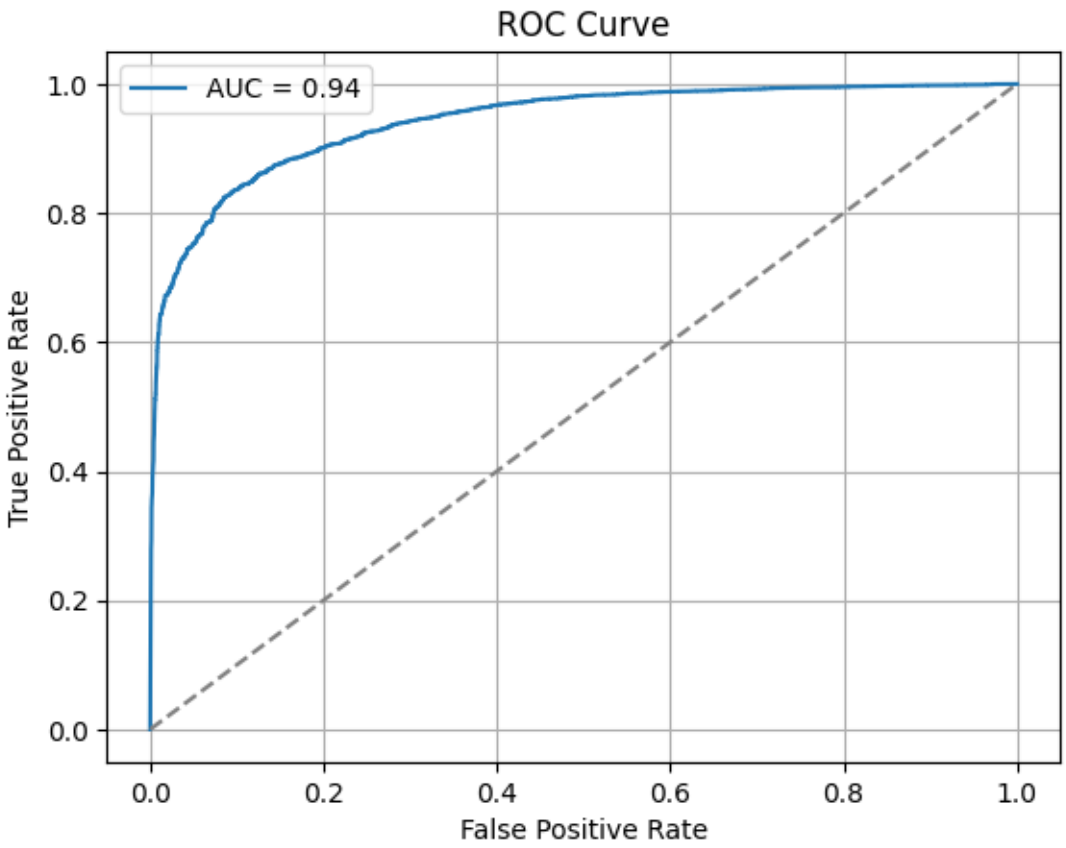
- **Test Accuracy**: 88.2%
- **Loss curves** showed a healthy convergence and low overfitting.
- **ROC AUC**: 0.94
- **Confusion Matrix**:




Model Accuracy
Accuracy
Epoch
Train Accuracy
Val Accuracy
0.85
0.80
0.75
0.70
0.65
0.60
0.55
0.50
0
5
10
15
20
25

-


Model Loss
Loss
Epoch
Train Loss
Val Loss
120
100
80
60
40
20
0
0
5
10
15
20
25



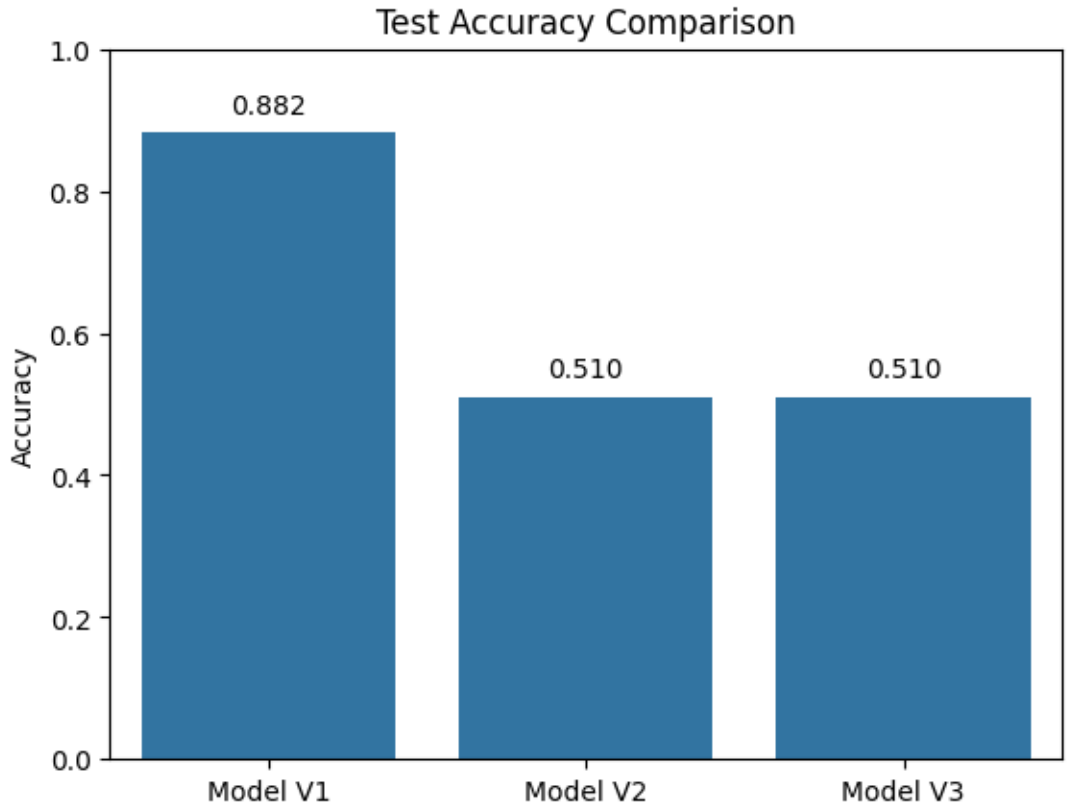
Model V2:

- **Test Accuracy**: 51.0%
- No meaningful learning observed. Accuracy stayed near chance-level (random guessing).
- Likely overfitting or instability due to deeper architecture without regularization.

Model V3:

- **Test Accuracy**: 51.0%
- Despite using L2 regularization and dropout, no performance gain observed.
- Training stabilized but converged to baseline performance.

Accuracy Comparison



Conclusion

Model V1 clearly outperformed the other two architectures. It achieved strong accuracy, converged well during training, and demonstrated good discriminative performance via its ROC curve. Model V2 and V3 did not improve upon the baseline and hovered around 51% test accuracy, suggesting no meaningful learning occurred in those versions.

Next Steps

- Fine-tune Model V1 further using hyperparameter tuning (e.g., learning rate, batch size)
- Explore feature engineering or richer input encoding (e.g., embeddings for categorical variables)
- Investigate class balance or possible label noise
- Test the model on real-world salary data or new samples for generalization performance