

# Predicting Loan Eligibility with PyTorch

Joseph Vazhiyil

## DATASET

The Loan Approval Classification Dataset is a synthetic dataset that was inspired from the Credit Risk Dataset. This dataset has 13 features that directly impact the loan approvals and one target variable named loan status. The dataset had 45k records, of which is mostly imbalanced. Most of the dataset is skewed towards the not approved status. The dataset was sourced from Kaggle.

## MODEL ARCHITECTURE

The model contains three fully connected layers. First layer has 13 input neurons and 64 output neurons. Then it's activated using ReLU. The second layer has 64 input neurons and 32 output neurons and is activated using ReLU. The second layer is the only hidden layer this model has. The third or output layer has 32 input neurons and 1 output neuron which is activated using sigmoid activation function.

## TRAINING PROCESS

### Model Training:

The train data is loaded, with a batch size of 32. Then we train this data over a loop of a hundred epochs. In each epoch, the gradients are set to zero, data is sent through the model. Loss calculated, and optimizer updates the weights.

### Hyperparameter Tuning:

The hyperparameters chosen to be augmented are loss functions, optimizers, learning rate and batch selections. We defined lists for all the hyperparameters, containing different values. We call a for loop for each list, making a nested for loop and redo the training process for each hyperparameter combination. The function will return the best combination with its test accuracy

## EVALUATION RESULTS

### Model Training Results:

Unseen data accuracy: 89%

|         | PRECISION | RECALL | F1   |
|---------|-----------|--------|------|
| Class 0 | 0.87      | 0.92   | 0.89 |
| Class 1 | 0.92      | 0.85   | 0.88 |

### Model Training Results After Fine Tuning:

Unseen Data accuracy: 90%

|         | PRECISION | RECALL | F1   |
|---------|-----------|--------|------|
| Class 0 | 0.90      | 0.90   | 0.90 |
| Class 1 | 0.90      | 0.90   | 0.90 |

## INSIGHTS

Overall, the model was performing well before optimizing it . It could correctly classify both classes with high accuracy indicating that it generalized the data well. For class 0 , the model seems to excel in identifying true negatives (0.87 precision) and classifying all true loan rejections (0.92). For class 1, recall is only 0.85 which means it's good at identifying loan approvals but still will miss actual approvals.

After hyperparameter tuning, the unseen data accuracy increased to 90 percent, indicating an improved generalization of the model. Also for both class 0 and class1, precision and recall became balanced. Recall decreased for class 0, whilst increased for class 1. This is a good tradeoff because now the model can identify loan rejections and approvals with a high level of accuracy. This is highly beneficial in a financial setting, as it improves the overall analysis of each loan application.