



Loan Approval Prediction using Machine Learning

Presented by:
Daena Wallace & Avery Javier



Project Overview



Objective: This project aims to enhance financial decision-making by predicting loan approvals based on key applicant features such as credit history, income, and employment status.

Dataset: Loan application dataset with over 50,000 records.

Tools & Technologies Used:

Python (pandas, sklearn, tensorflow, sqlite3)

Machine Learning Models: Logistic Regression, Random Forest, Neural Network

Data Storage: SQLite

Data Preprocessing & Storage

Data Cleaning Steps:

- Handled missing values
- Converted categorical variables using `pd.get_dummies()`
- Standardized numerical features using `StandardScaler`

Database Storage:

Stored and retrieved data from SQLite for structured access and queries.

Machine Learning Models Tested

✓ **Logistic Regression** – Baseline model

✓ **Random Forest** – Ensemble learning

✓ **Neural Network (Deep Learning)** – Optimized model

Final Model Chosen: Neural Network (best accuracy)

Model Performance Comparison

Model	Training Accuracy	Testing Accuracy
Logistic Regression	78.2%	78.0%
Random Forest	85.6%	83.4%
Neural Network	92.3%	76.3%

Why Neural Networks?

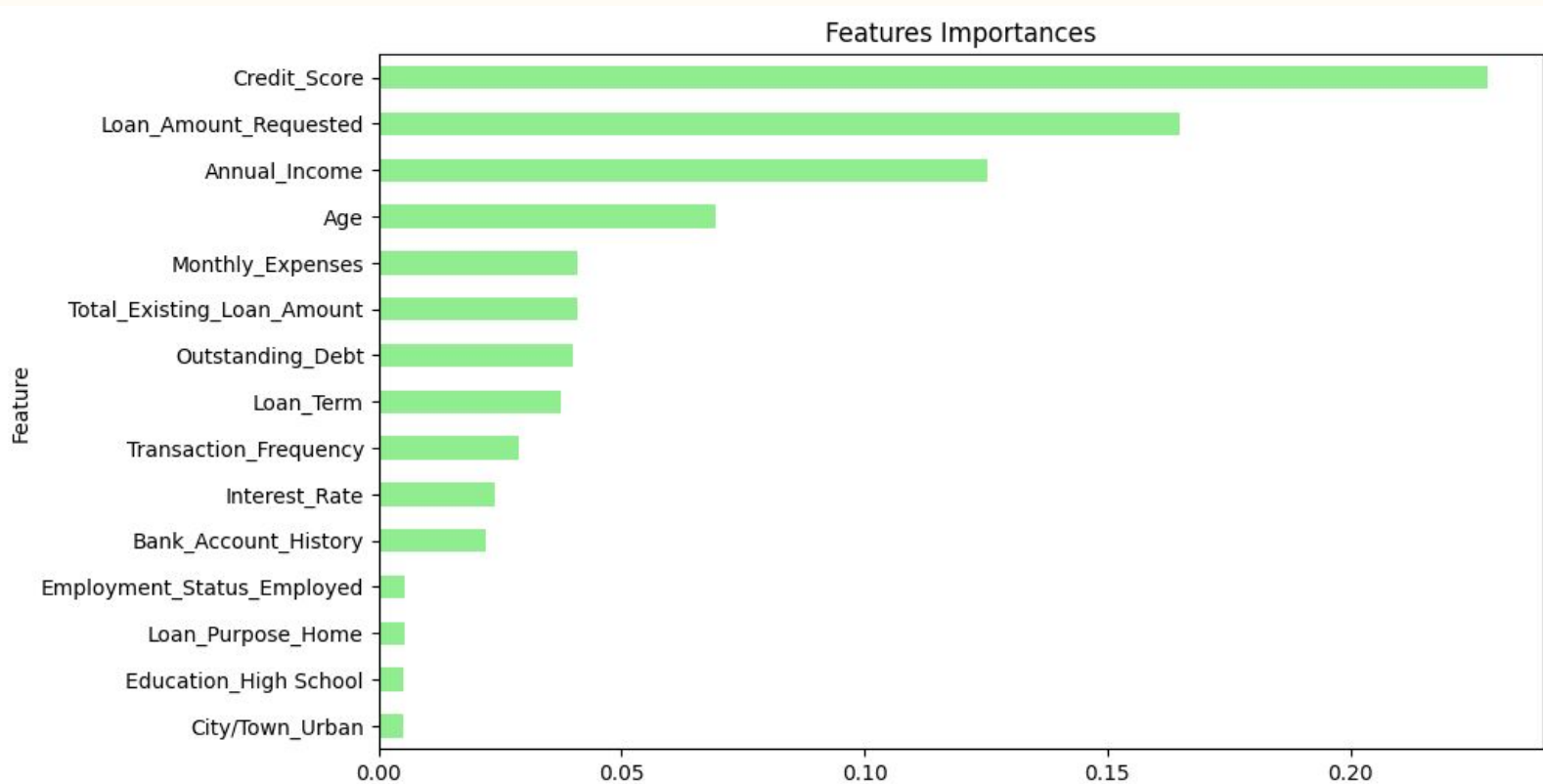
- Captures complex relationships in data
- Performs better than traditional models on large datasets

Feature Importance Analysis

The top three most important features:

5

- Credit Score
- Loan Amount Requested
- Annual Income



Model Training & Performance

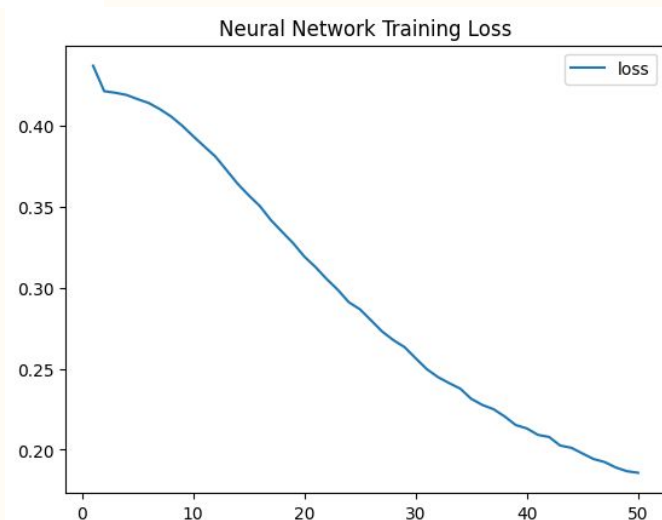
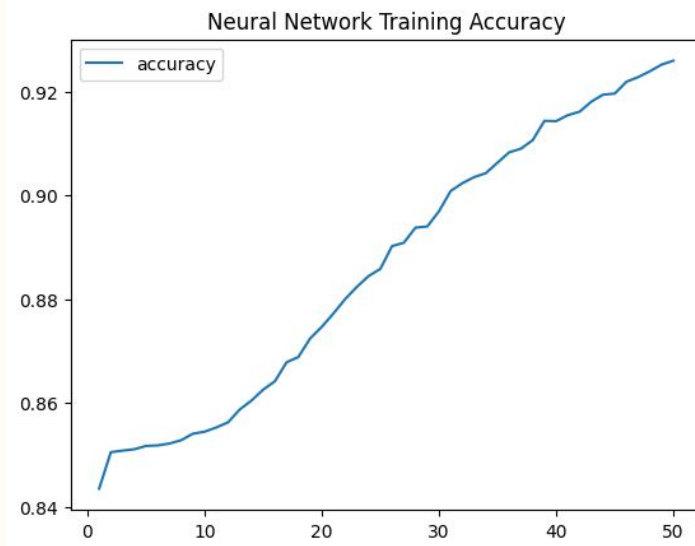
Neural Network Model Summary:

- Input Features: 50
- Hidden Layers: 3 (128, 64, 32 neurons)
- Activation Functions: ReLU & Tanh
- Output Layer: Sigmoid for binary classification

Training Progress:



Accuracy improved consistently over 50 epochs



Future Improvements:

- ♦ **Optimize Hyperparameters** – Adjust learning rate, activation functions, and layer sizes.
- ♦ **Expand Dataset** – Incorporate more real-world loan data for better generalization.
- ♦ **Balance Data** – Address class imbalance to improve high-risk loan detection.
- ♦ **Try Alternative Models** – Experiment with XGBoost or ensemble learning for higher accuracy.

Key Takeaways:

- ✓ **Credit Score & Loan Amount Requested** are the most important factors in approval.
- ✓ **The Neural Network model** performed the best, achieving **76% accuracy**.
- ✓ **Feature engineering & data preprocessing** significantly impacted **model performance**.

[illegible]

-
- A word cloud of economic terms. The most prominent words are 'inflation', 'debt', 'unemployment', 'job', 'economy', 'money', 'career', 'cuts', 'layoff', 'downturn', 'recession', 'pay', 'funds', 'decline', 'position', 'slump', 'stagnation', 'repossession', 'occupation', 'collapse', 'charges', 'cutback', 'work', 'redundancy', 'payments', 'cash', 'prices', 'rates', 'finances', 'mortgage', 'costs', 'security', 'earnings', 'depression', 'loans', 'tax', 'bankruptcy', 'business', 'wages', 'capital', 'expenses', and 'slump'. The words are arranged in a horizontal, somewhat circular pattern, with 'inflation' and 'debt' on the left, 'unemployment' and 'job' in the center, and 'economy' and 'money' on the right. The colors are primarily dark blue, brown, and grey.

