SUBMITTED BY GROUP 4

- Kimberly Wessler
- Cris Pena
- Jajuan Graham
- Chris Dobbin
- Christine Espiritu

📊 Project Overview

This project investigates the effectiveness of machine learning algorithms in predicting credit risk using dataset from Kaggle. The goal is to build a model that improves predictive accuracy while maintaining transparency and fairness, thereby supporting responsible and inclusive lending decisions.

Workflow

📥 **Data Ingestion**

🧪 **Data Preprocessing**

📊 **Data Analysis**
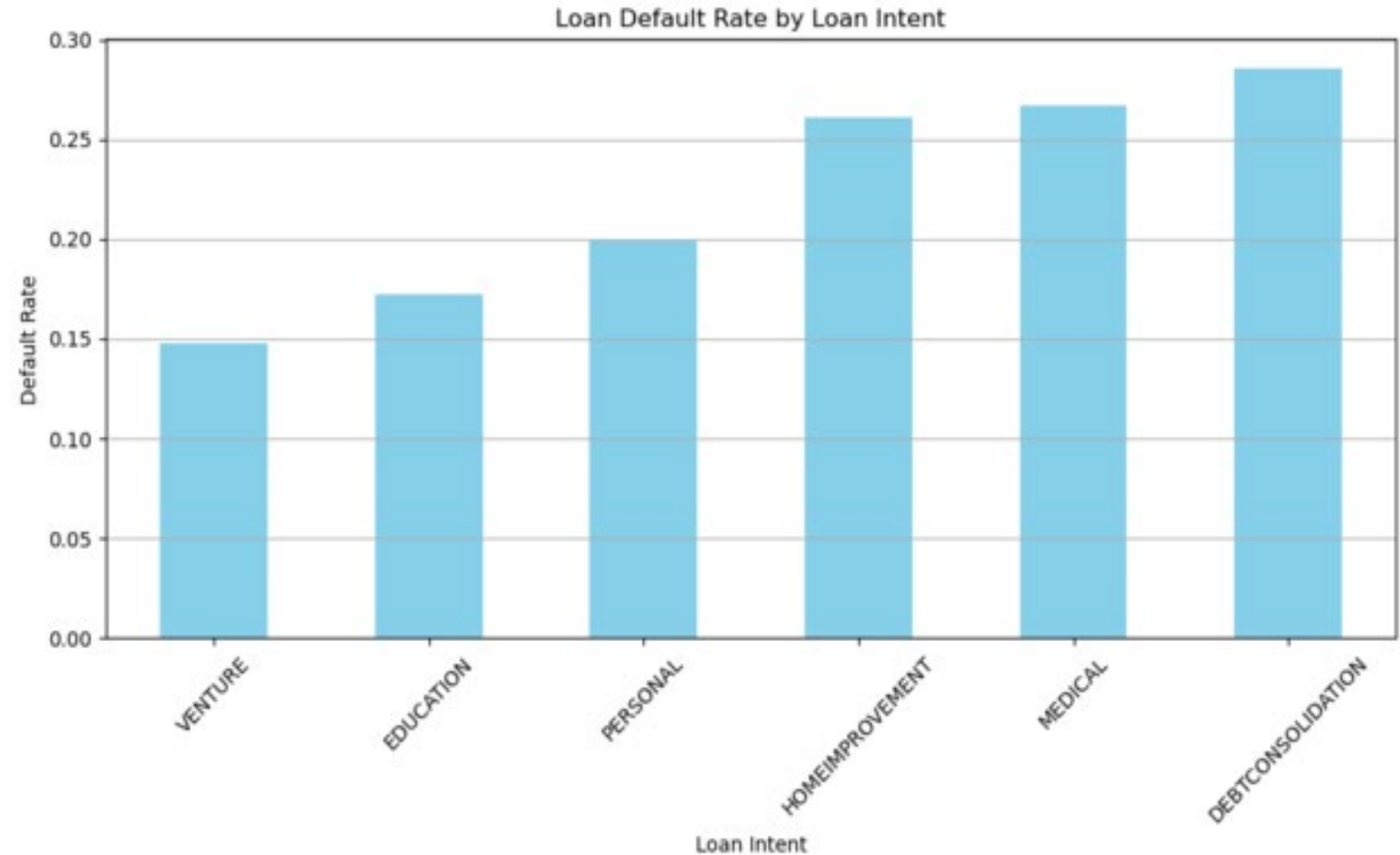
🔀 **Nueral Network Model**

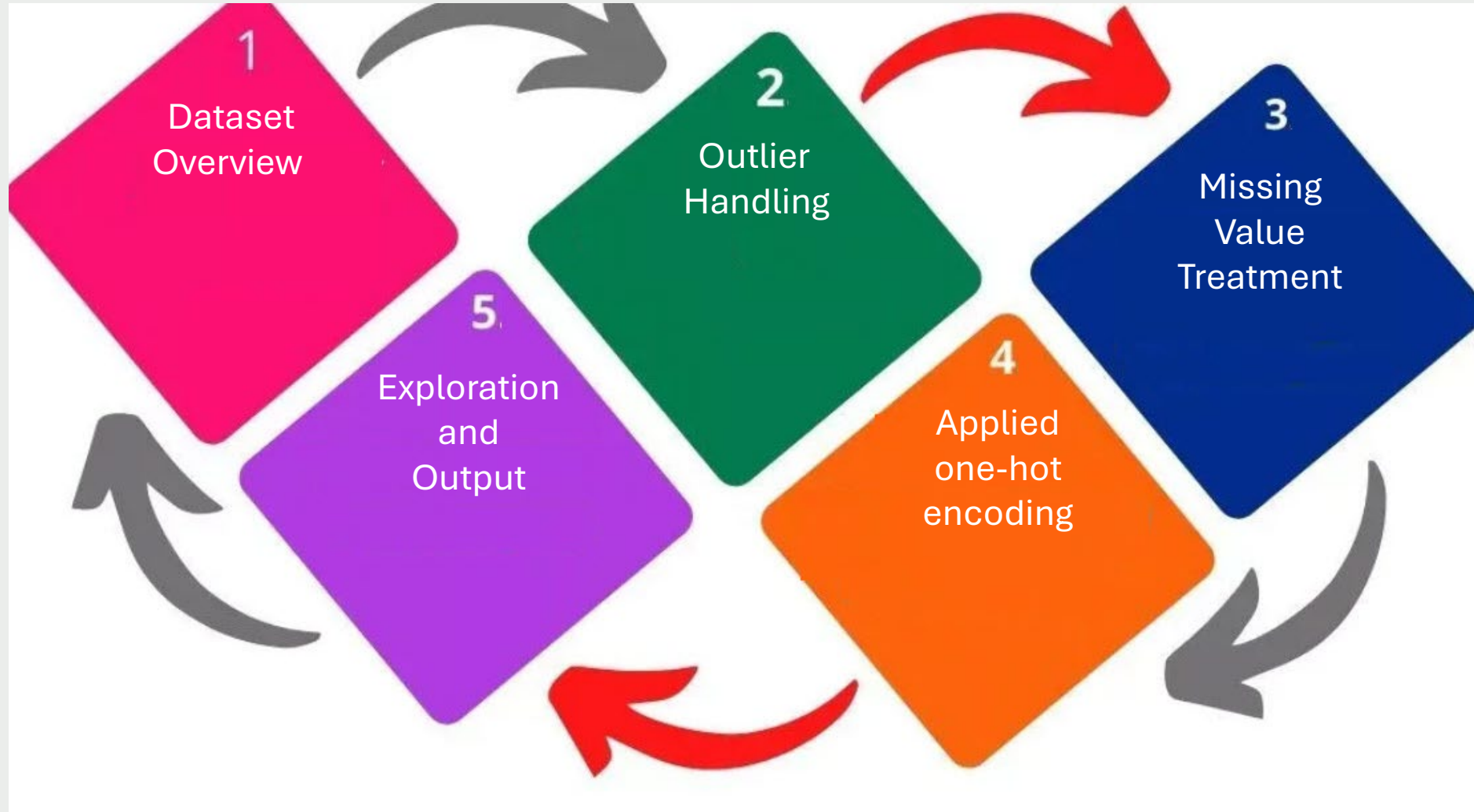🌲 **Random Forest Model Training**
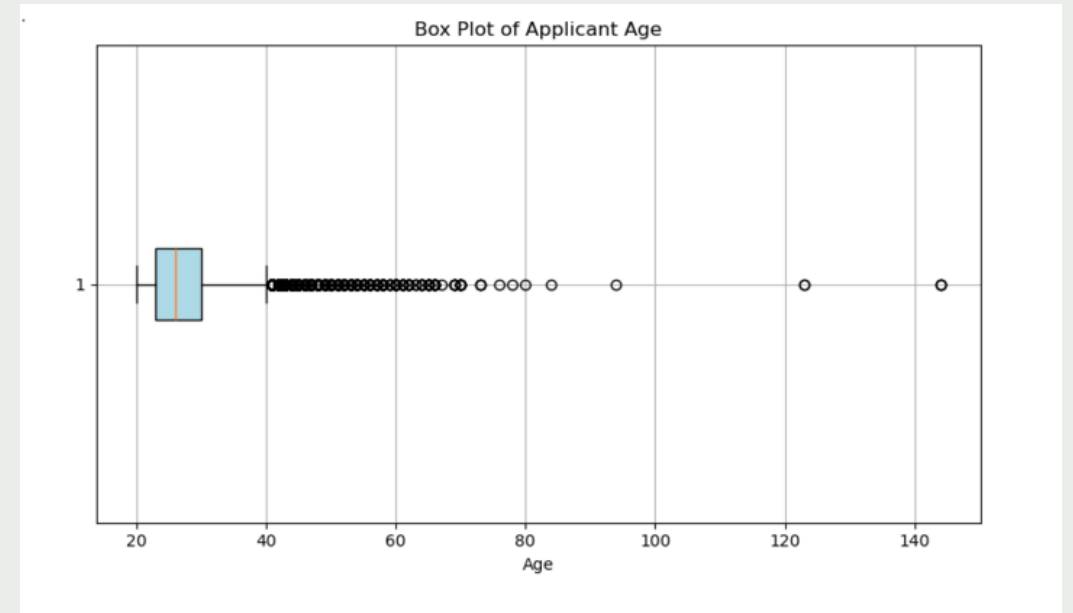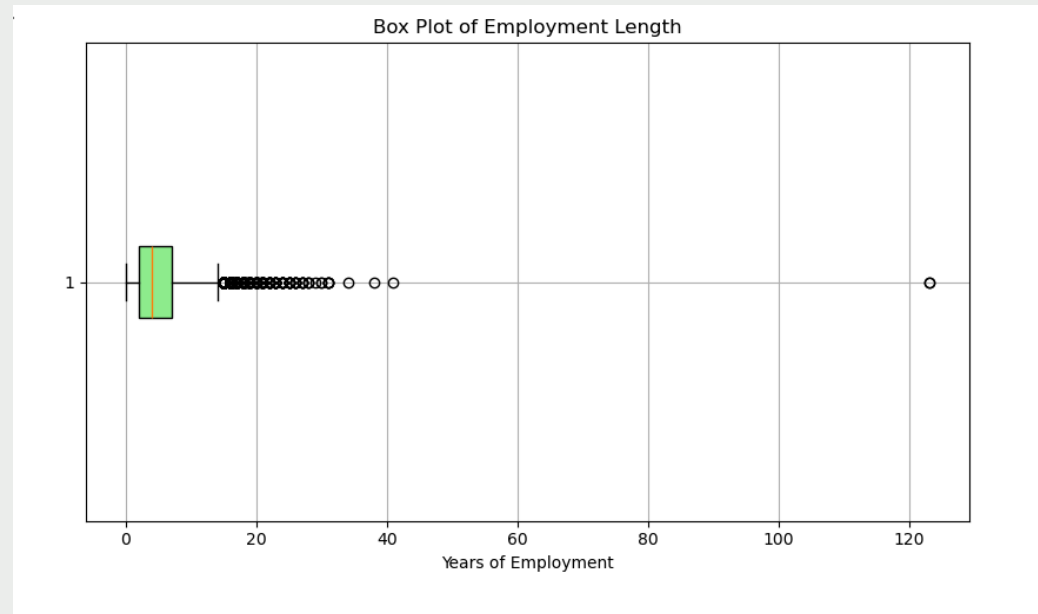
🧪 **Model Evaluation**

Key Factors

- Age
- Income
- Home Ownership
- Length of Employment
- Loan Intent
- Loan Grade
- Loan Amount
- Loan Interest Rate
- Loan Status
- Loan Percent Income
- Default on file
- Length of Credit History

Loan Default Rate by Loan Intent

Data Preprocessing Pipeline

Outliers

Model Development

Model 1: Baseline Model

Used the **Adam** optimizer

Simple architecture with 1–2 dense layers

Each Layer had 16 –lowed for **deeper networks** with 2–4 layers and up to **256 units** per layer. 128 units

Model 2: Enhanced Model

Introduced **AUC** as a secondary metric to better handle potential class imbalance.

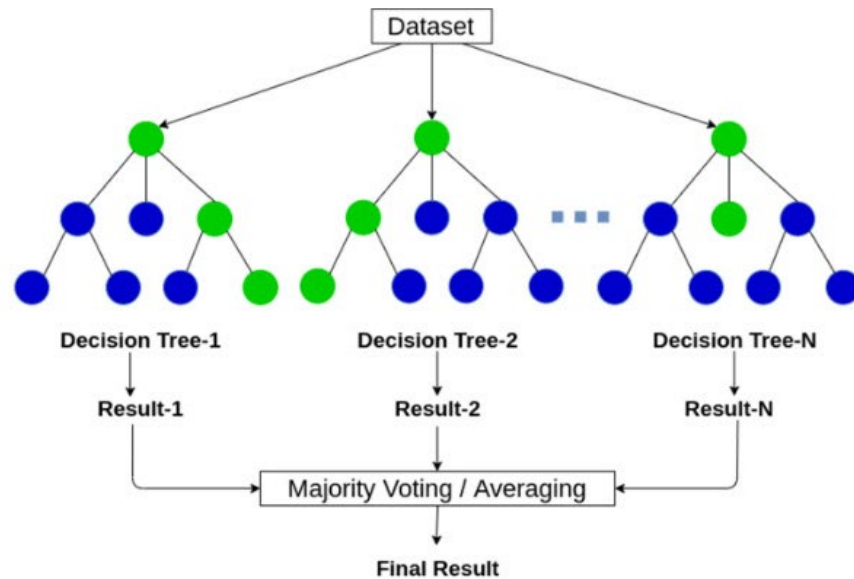Has **deeper networks** with 2–4 layers and up to **256 units** per layer

Model Performance

Both models achieved **85% accuracy** on test data.

Model 1 lacked overfitting protection and thus had lower generalizability.

**Model 2** incorporated regularization and early stopping, making it more robust and production-ready.
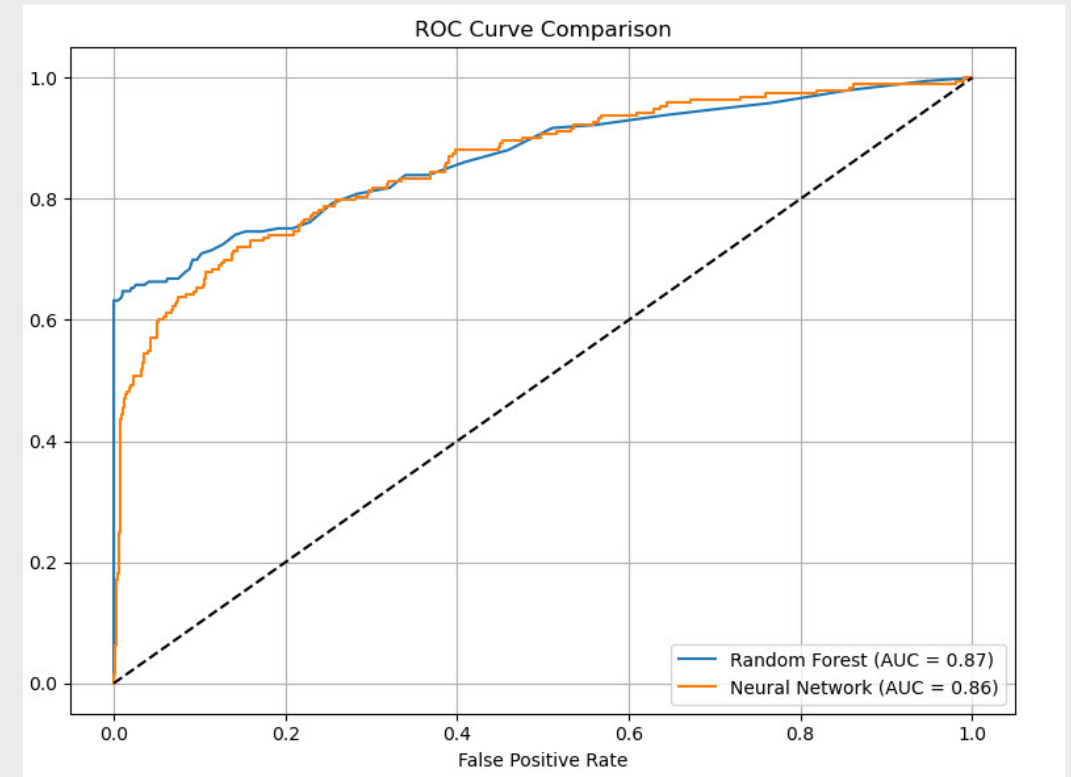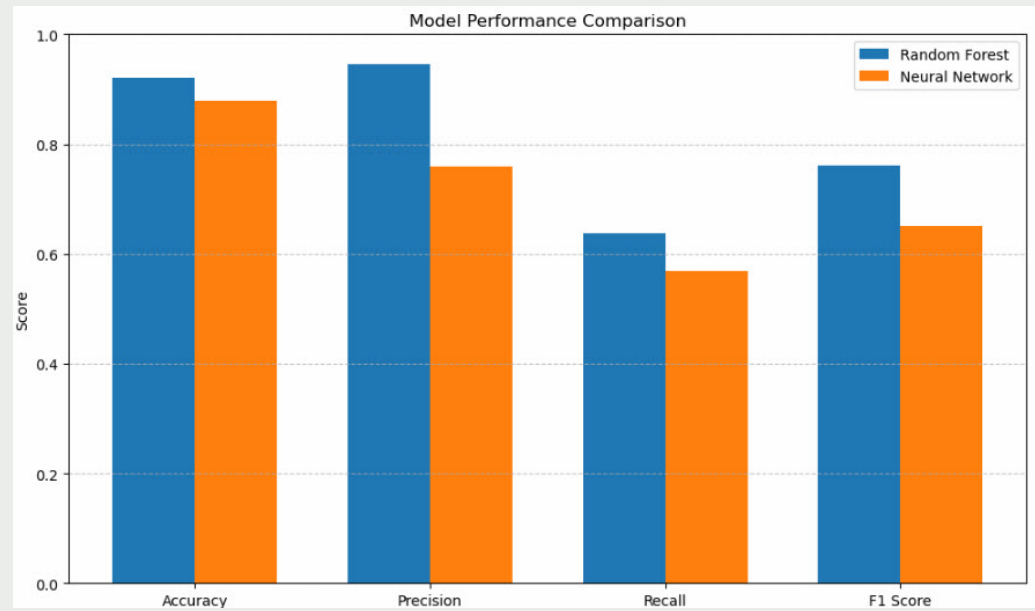
Random Forest

Random Forest Model Evaluation	Test 1	Test 2	Test 3
Accuracy	92.01%	92.32%	92.62%
Precision	91.89%	94.07%	98.59%
Recall	62.50%	65.46%	66.67%
F1 Score	74.40%	77.20%	79.55%

Random Forest Model Evaluation	Results
Accuracy	89.96%
Precision	86.84%
Recall	62.86%
F1 Score	72.93%

Modified Random Forest Modeling
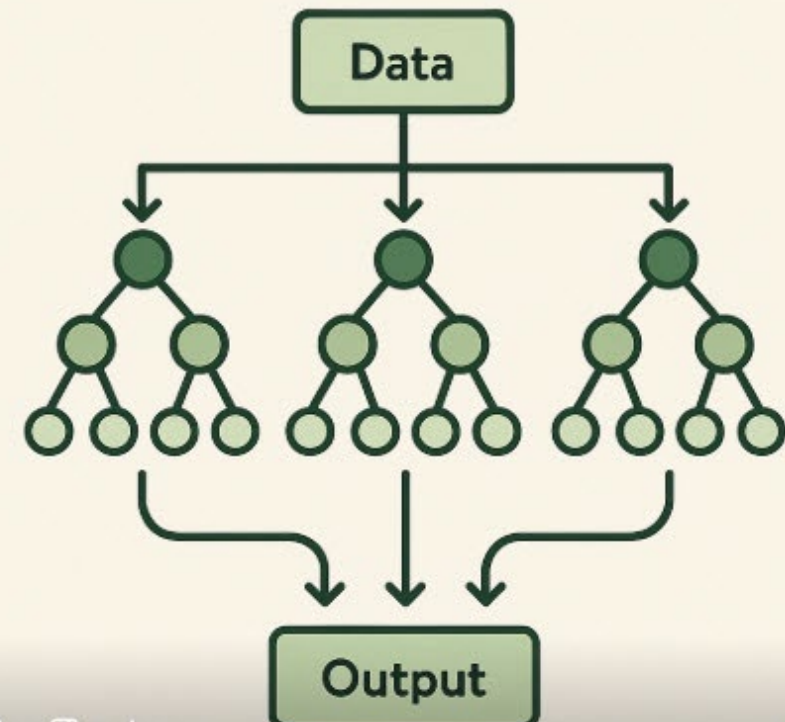
Model Comparison Graph

THANK YOU

- Kimberly Wessler
- Cris Pena
- Jajuan Graham
- Chris Dobbin
- Christine Espiritu