Final Project Report: AI Model Comparison

Course: Introduction to AI  
Project Title: Comparative Analysis of Five AI Models  
Submitted by: Hunter Africk, Joey Paschke  
Date: 12/12/24

**1. Introduction and Objective (5 Marks)**

**Objective:**Compare the performance of multiple AI models in predicting the likelihood of success in the music industry given the user's country and preferred genres of music.

**Problem Statement:**The task will involve predicting how successful a user may be in the music industry, given their country and three chosen genres of music. Models are trained on a dataset of music artists, their popularity, and associated attributes to make predictions.

**Overview of AI Models Chosen:**

| Model Number | Model Name | Purpose |
| --- | --- | --- |
| 1 | Logistic Regression | Get a baseline for predicting success likelihood. |
| 2 | Random Forest | Capture the non linear relationships in listener counts and country. |
| 3 | Gradient Boosting (XGBoost) | Brief purpose of Model 3 |
| 4 | Neural Networks | Brief purpose of Model 4 |
| 5 | Support Vector Machines | Brief purpose of Model 5 |

**2. Justification of Model selection (2 Marks)**

Justification for Model Selection:

| Model Name | Reason for Selection |
| --- | --- |
| Logistic Regression | It's a simple model |
| Random Forest | Justification |
| Gradient Boosting (XGBoost) | Justification |
| Neural Networks | Justification |
| Support Vector Machines | Justification |

**3. Model Descriptions (1 Marks)**

Model Overview:

| Model Number | Model Name | Architecture Details | Key Features |
| --- | --- | --- | --- |
| 1 | Logistic Regression | Architecture details | Key features of Model 1 |
| 2 | Random Forest | Architecture details | Key features of Model 2 |
| 3 | Gradient Boosting | Architecture details | Key features of Model 3 |
| 4 | Neural Networks | Architecture details | Key features of Model 4 |
| 5 | Support Vector Machines | Architecture details | Key features of Model 5 |

**4. Dataset Description (2 Marks)**

Dataset Information:

| Dataset Attribute | Description |
| --- | --- |
| Name | Music Artists Popularity |
| Source | <https://www.kaggle.com/datasets/pieca111/music-artists-popularity/data> |
| Size | 118,000 Unique Artists, |
| Class Distribution |  |
| Preprocessing Steps | Remove unwanted columns, remove duplicate rows, remove rows with missing values, removed rows with at least two `0.0` values. |

Example

| Name | CIFAR-10 |
| --- | --- |
| Source | Available from <https://www.cs.toronto.edu/~kriz/cifar.html> |
| Size | 60,000 images (50,000 for training, 10,000 for testing) |
| Class Distribution | 10 classes, with 6,000 images per class |
| Preprocessing Steps | Normalization to range [0, 1], data augmentation (random cropping, horizontal flipping) |

**Dataset Justification:**  
Explain why this dataset is suitable for the models.

**5. Experimental Setup (10 Marks)**

Experimental Design: If any other metric is used, add it to the table

| Metric |
| --- |
| Accuracy |
| Precision |
| Recall |
| F1 Score |

Parameter Settings:

| Model Name | Hyperparameter 1 | Hyperparameter 2 | Hyperparameter 3 | Hyperparameter 4 | Additional Notes |
| --- | --- | --- | --- | --- | --- |
| Model 1 | Value | Value | Value | Value | Any other notes |
| Model 2 | Value | Value | Value | Value | Any other notes |
| Model 3 | Value | Value | Value | Value | Any other notes |
| Model 4 | Value | Value | Value | Value | Any other notes |
| Model 5 | Value | Value | Value | Value | Any other notes |

Environment Details:

| Component | Specification |
| --- | --- |
| Operating System | Windows |
| Software Version | Windows 11 |
| Hardware | Personal Computer |
| Link to the code base |  |

**6. Results and Analysis (50 Marks)**

Performance Metrics:

| Model Name | Accuracy (%) | Precision (%) | Recall (%) | F1 Score (%) | Additional Metrics |
| --- | --- | --- | --- | --- | --- |
| Model 1 | [Value] | [Value] | [Value] | [Value] | [Value] |
| Model 2 | [Value] | [Value] | [Value] | [Value] | [Value] |
| Model 3 | [Value] | [Value] | [Value] | [Value] | [Value] |
| Model 4 | [Value] | [Value] | [Value] | [Value] | [Value] |
| Model 5 | [Value] | [Value] | [Value] | [Value] | [Value] |

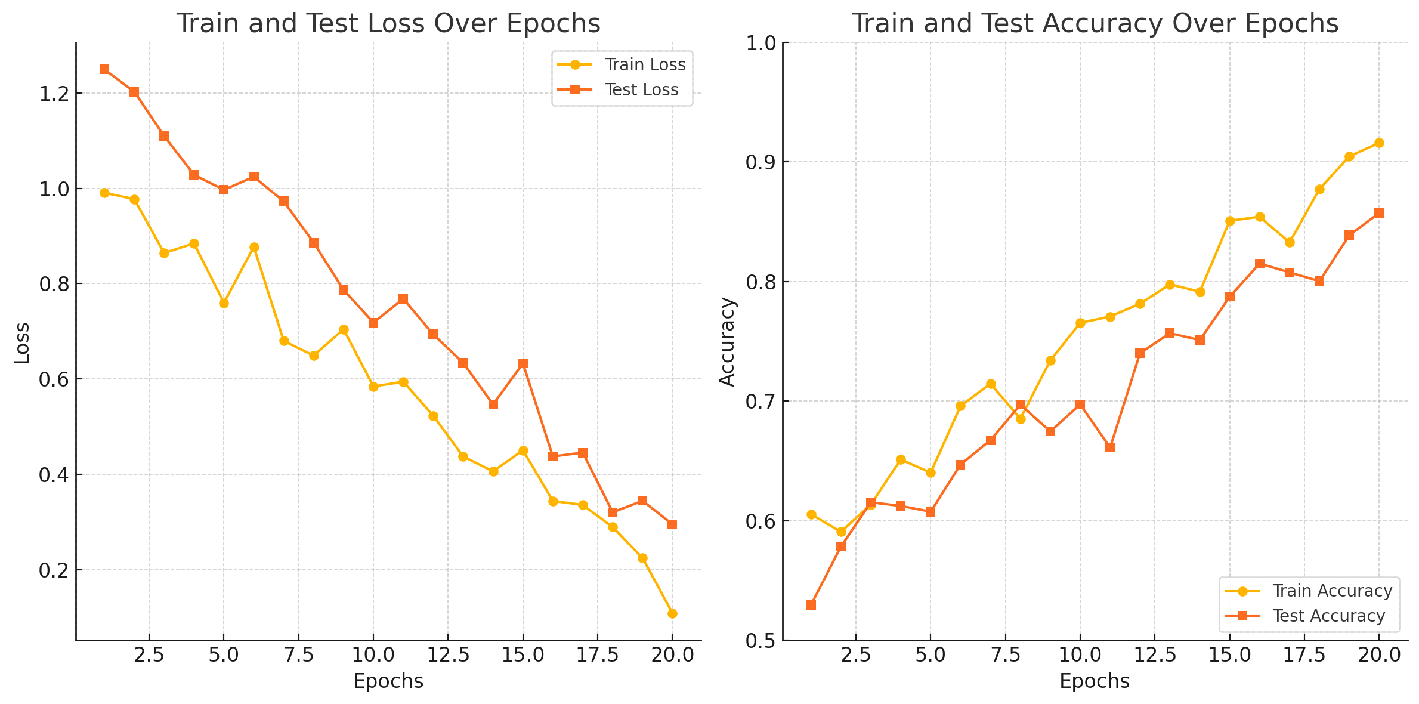
Comparative Analysis:  
Summarize and analyze differences in performance.

Error Analysis:  
Identify main sources of error and discuss possible reasons.

**Plots**

**- Training and Test accuracies**

**For neural network-based model, plot the following**



**7. Discussion and Insights (10 Marks)**

Interpretation of Results:  
Discuss key insights gained from the model comparison.

Limitations:  
Mention any limitations in the experimental design, data, or model performance.

Future Directions:  
Provide suggestions for further improvements or research.

**8. Conclusion (5 Marks)**

Summarize the project findings and the strengths and weaknesses of each model.