AI Art Advisor: A Machine Learning Approach to Art Style Classification

Heeje Yoo – August 13, 2025

This paper documents the methodology and findings for the “AI Art Advisor,” a classification system designed to identify the style of an artwork from a digital image. It evaluates an ensemble of machine learning algorithms to determine the most effective model and discusses its application in improving art accessibility and education.

Methodology

The system follows a full data mining workflow, from acquisition and preprocessing to feature extraction and model evaluation.

Data Sourcing and Preparation – The “WikiArt All Artpieces” dataset (Kaggle) was used, focusing on 15 prominent art styles (e.g., Impressionism, Cubism, Surrealism). To avoid class imbalance, a maximum of 2,000 images per style was sampled, producing 30,000 images. Corrupted files were removed, leaving 29,874 valid images.

Feature Extraction with Transfer Learning – A pre-trained EfficientNetB0 CNN (trained on ImageNet) served as a feature extractor. Images were resized to 224×224 pixels, passed through the network (top layer removed, GlobalAveragePooling2D added), and converted into 1,280-dimensional vectors capturing high-level patterns and textures. This leveraged deep learning without the cost of training a large CNN from scratch.

Model Training and Evaluation – Features were split into training (70%) and test (30%) sets, standardized, and fed into seven algorithms: Logistic Regression, Gaussian Naive Bayes, Decision Tree, Random Forest, K-Nearest Neighbors, K-Means, and SVM. Each was evaluated with 5-fold cross-validation; the SVM underwent GridSearchCV tuning ({'C': 50, 'gamma': 0.001, 'kernel': 'rbf'}). Performance was assessed via accuracy, precision, recall, and F1-score.

Findings and Model Selection

The optimized SVM outperformed all others, with 62.2% cross-validation and 63.7% test accuracy. Decision Trees performed poorly (23.5%), Random Forests moderately (44.9%), and KNN/Logistic Regression achieved ~50% accuracy. The SVM’s ability to handle high-dimensional, non-linear relationships made it ideal for this problem, leading to its selection for deployment.

Real-World Application

The “AI Art Advisor” can enhance art appreciation by enabling instant style identification and contextual explanations. Potential uses include:

- **Educational Tool** – Supports art history learning with immediate style recognition.
- **Museum/Gallery Companion** – Enhances visitor engagement through mobile style identification.
- **Recommendation Engines** – Powers style-based suggestions in online galleries or marketplaces.

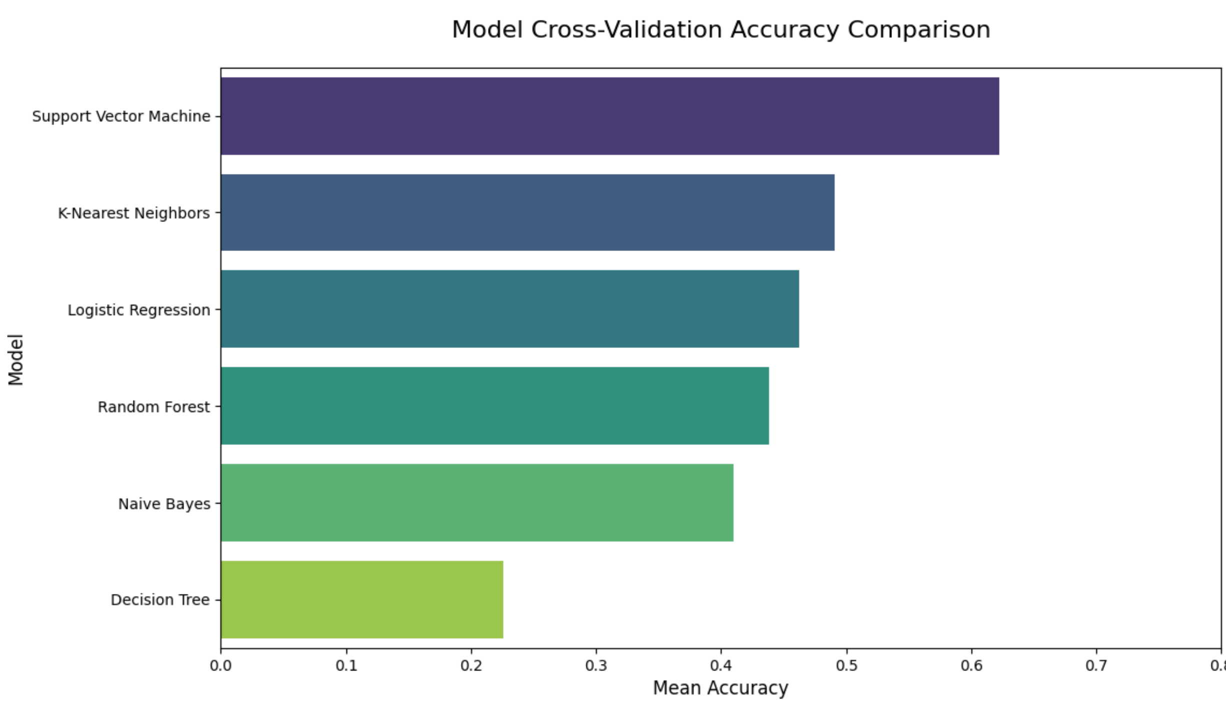
By automating art style classification, this project lays the groundwork for tools that enrich cultural education and foster broader appreciation of the arts.

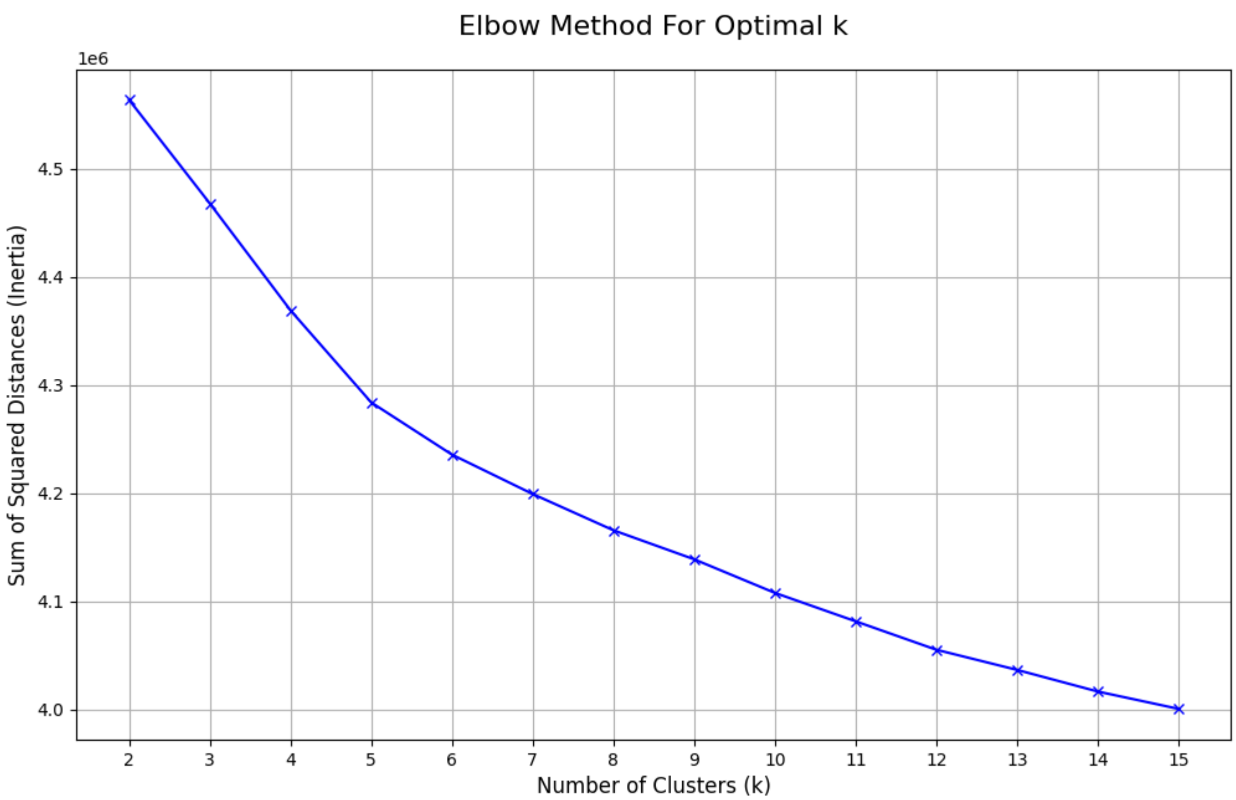
Appendix

--- Model Comparison ---

	Model	Cross-Validation Accuracy	Test Accuracy \
0	Logistic Regression	0.462756	0.480701
1	Naive Bayes	0.409590	0.414770
2	Decision Tree	0.226190	0.234828
3	Random Forest	0.438612	0.449018
4	Support Vector Machine	0.622490	0.636881
5	K-Nearest Neighbors	0.491012	0.496542

	Precision (Weighted Avg)	Recall (Weighted Avg)	F1-Score (Weighted Avg)
0	0.475966	0.480701	0.477955
1	0.408554	0.414770	0.405322
...			
4	0.636043	0.636881	0.635697
5	0.519051	0.496542	0.494704





References

Lopes, S. (2022). *WikiArt All Artpieces*. Kaggle. Retrieved August 13, 2025, from https://www.kaggle.com/datasets/simolopes/wikiart-all-artpieces