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 **29,907** valid images for the analysis.
- *Feature Extraction with Transfer Learning:* A pre-trained **EfficientNetB7** 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 **2,560-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 six classification algorithms: Logistic Regression, Gaussian Naive Bayes, Decision Tree, Random Forest, K-Nearest Neighbors, and a Support Vector Machine (SVM). A separate K-Means cluster analysis was also performed to explore the natural groupings within the feature data.
  Each classification model was evaluated with 5-fold cross-validation; the SVM underwent GridSearchCV tuning, which found the optimal parameters to be **{'C': 10, 'gamma': 0.0001, 'kernel': 'rbf'}**. Performance was assessed via accuracy, precision, recall, and F1-score.

Findings and Model Selection
The optimized **SVM** outperformed all other models, though overall accuracy was lower than initially reported. It achieved a **47.4% cross-validation accuracy** and a **48.0% test accuracy**. The SVM's superior ability to handle the high-dimensional, non-linear relationships in the feature space made it the best choice for this problem, leading to its selection for deployment.

Real-World Application
The "AI Art Advisor" enhances art appreciation with instant style identification and explanations, serving as an educational tool, museum companion, or recommendation engine. By automating art style classification, this project provides a foundation for tools that enrich cultural education and foster a broader appreciation of the arts.
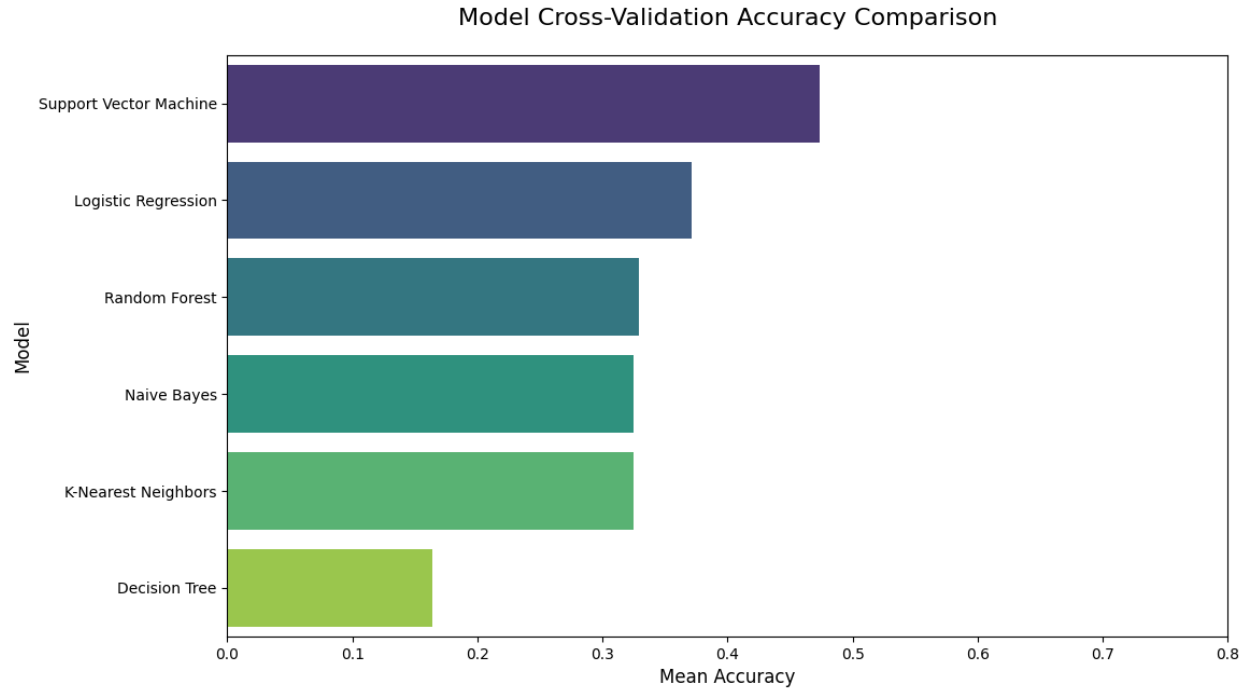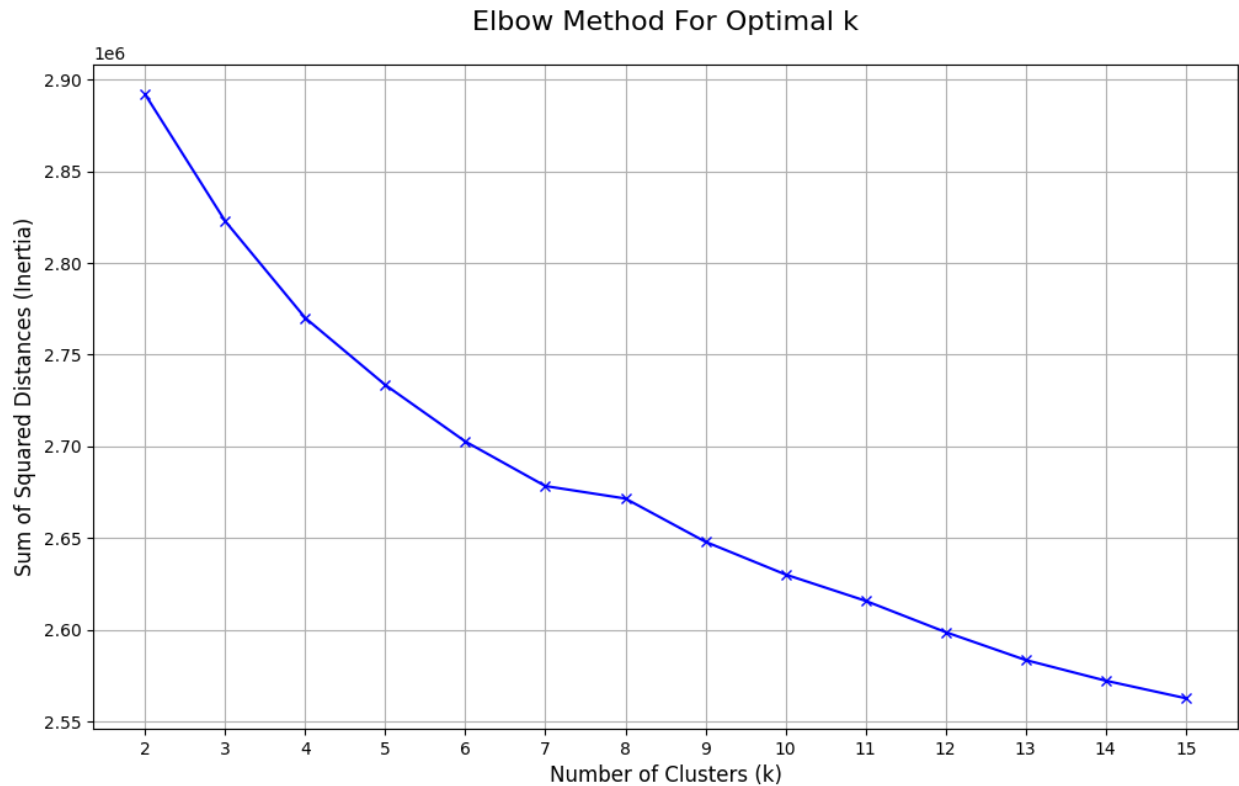
Appendix

Model Comparison

Model	Cross-Validation Accuracy	Test Accuracy	Precision (Weighted Avg)	Recall (Weighted Avg)	F1-Score (Weighted Avg)
Support Vector Machine	**47.4%**	**48.0%**	**47.7%**	**48.0%**	**47.7%**
Logistic Regression	37.1%	36.9%	37.1%	36.9%	37.0%
Random Forest	32.9%	33.9%	32.8%	33.9%	32.5%
K-Nearest Neighbors	32.5%	33.1%	33.5%	33.1%	31.9%
Naive Bayes	32.5%	32.7%	33.3%	32.7%	31.7%
Decision Tree	16.4%	16.4%	16.6%	16.4%	16.5%

```
--- Model Comparison ---
                 Model  Cross-Validation Accuracy  Test Accuracy  \
0    Logistic Regression                   0.371118       0.368773
1           Naive Bayes                   0.324830       0.326647
2         Decision Tree                   0.164326       0.164493
3         Random Forest                   0.328843       0.338794
4  Support Vector Machine                   0.473774       0.479996
5     K-Nearest Neighbors                   0.324591       0.330882

   Precision (Weighted Avg)  Recall (Weighted Avg)  F1-Score (Weighted Avg)
0                  0.371414               0.368773                 0.369812
1                  0.333425               0.326647                 0.317255
2                  0.165552               0.164493                 0.164899
3                  0.327856               0.338794                 0.325242
4                  0.476960               0.479996                 0.476894
5                  0.335384               0.330882                 0.318501
```

Model Cross-Validation Accuracy Comparison

Model Performance Metrics on Test Data

Elbow Method For Optimal k


```
--- Analyzing clusters with k=7 ---
Cross-tabulation of True Art Styles vs. K-Means Cluster ID:
Cluster_ID                  0     1     2     3     4     5     6
True_Style
Abstract Expressionism     17    82  1306   126   443    24     2
Art Nouveau (Modern)      134   726   347   127   298   191   173
Baroque                   126   228    73   189   252   681   447
Cubism                     61   686   685    57   382    98    27
Expressionism              92   595   449   148   304   208   199
Impressionism             248   147   101   767   339   173   218
Neoclassicism             303   280   115   136   407   237   510
Northern Renaissance      135   350   132    94   290   592   388
Pop Art                    50   427   679    32   652    50   110
Post-Impressionism        310   377   224   513   248   148   173
Realism                   115   281    78   529   370   212   408
Rococo                    143   127    58   119   183   466   893
Romanticism               161   208    34   569   275   374   371
Surrealism                 86   368   552   129   662   136    64
Symbolism                 143   364   301   521   273   251   145
```

References

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