Springboard Data Science Course
Capstone Project 3

Classification of Real and AI-Generated Images using Unsupervised and Supervised Models

By: Priyanka Panhalkar

March, 2025

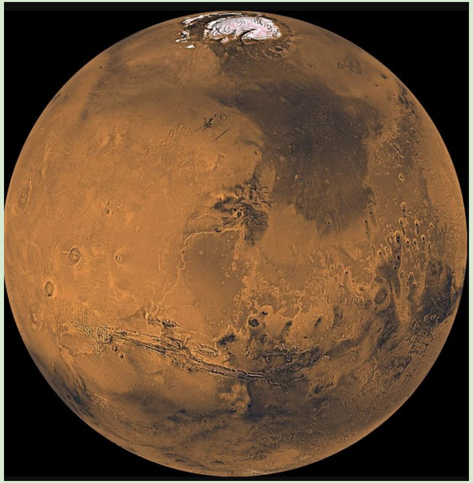
Introduction

Business Problem: With the rise of Generative AI (GenAI), distinguishing between real and AI-generated images is challenging. We aim to develop an AI model that automatically detects whether an image is real or AI-generated, improving trust, security, and transparency across industries

Approach : We implemented a cluster-based classification approach by applying KMeans on L2-normalized CLIP embeddings, using these embeddings as features to assign cluster labels via majority voting from the training set. Test samples were then classified based on proximity to cluster centroids. In parallel, we trained supervised models like **Logistic Regression, Random Forest**, and **Support Vector Machine (SVM)**, which directly learned decision boundaries using these embeddings. SVM achieved the highest performance, with F1 score ~92% across all metrics

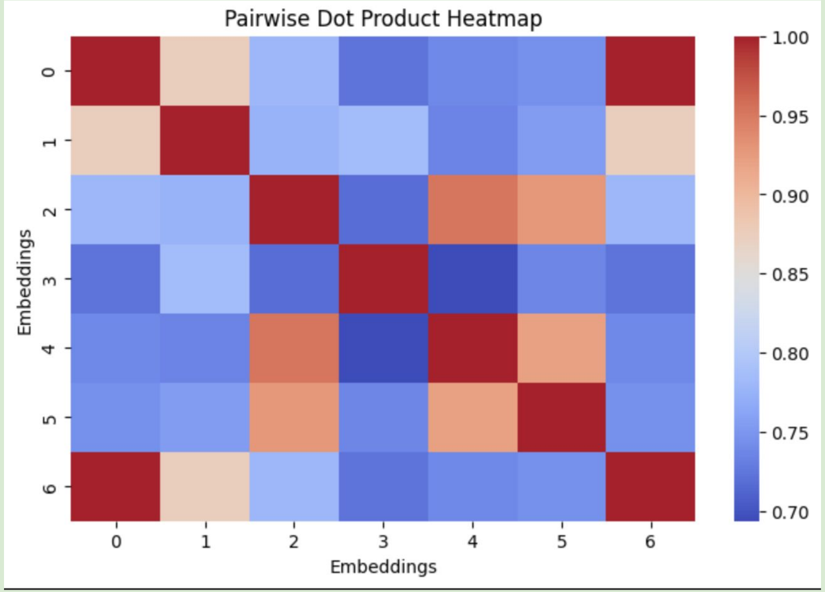
Data Collection

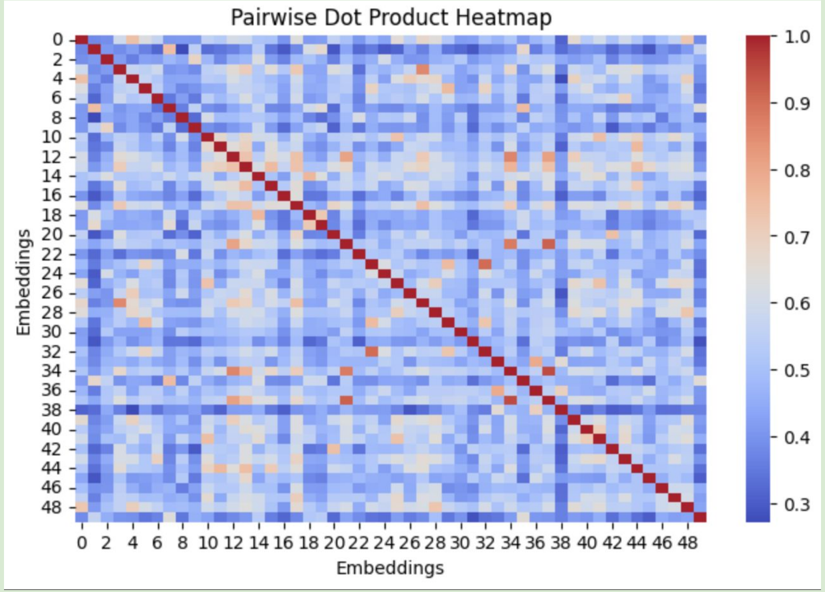
Data Collection: The Real and Gen AI images were gathered from Pinterest website for multiple categories like Statue of Liberty, Cities, Mars etc. 3090 total images were collected. Each image was converted to a 512 dimensions embedding using the CLIP **transformer**, an OpenAI Convolutional Neural Network (CNN) Model. Sample Images.eg. First 1,2 are Gen AI 3,4 are Real



Exploring and Data Analysis (1)

Multiple dot product heatmaps were generated using normalized embeddings to explore the similarity patterns within AI-generated images, real images, and between the two groups. Eg. First image- Mars images of Real and GenAI. Second Image Random 50 mixed images

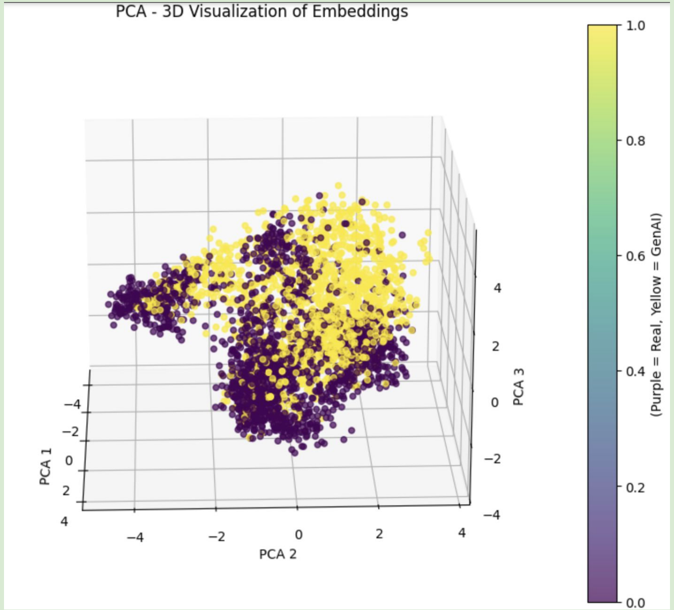


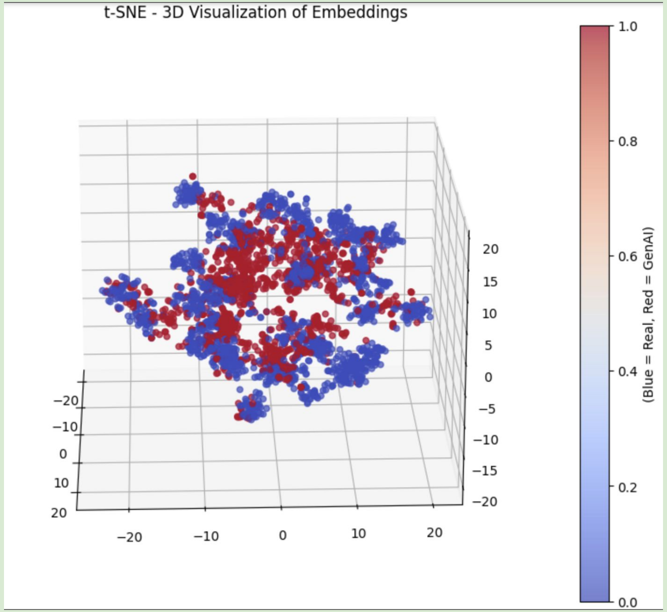


Exploring and Data Analysis (2)

Dimensionality reduction using PCA (2D and 3D) revealed distinct clustering patterns in the embedding space, indicating separation between real and AI-generated images.

2D and 3D t-SNE visualizations further emphasized local structure and neighborhood similarities, reinforcing the suitability of these embeddings for downstream classification tasks.

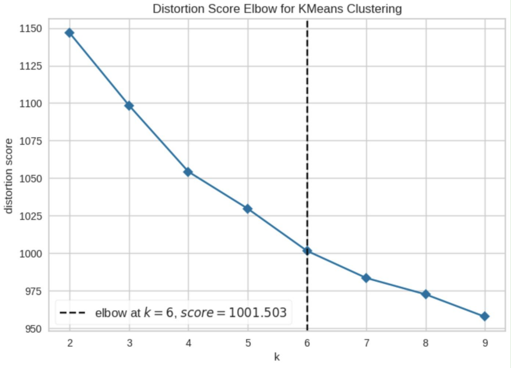


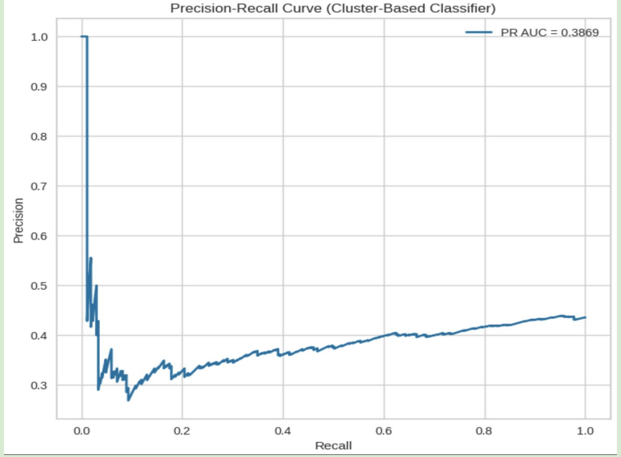


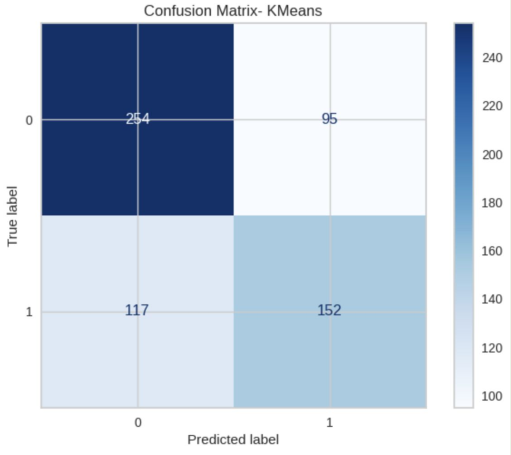
Results and Visualizations (Unsupervised Modeling)

Cluster Based Classification: Goal was to understand structure without using labels

- Clustering with KMeans: Using Elbow Method where K=6
- Distance to Centroid Labeling: Test images were assigned to the nearest cluster centroid and labels were predicted via Majority Voting.
- Results: **Confusion Matrix**: 254 TN, 152 TP, 117 FN , 95FP
- Insights from Matrix: Real images (0) exhibit more dispersion and variability while AI images cluster tightly, indicating homogeneity
- This validates that embeddings are informative and support downstream classification tasks







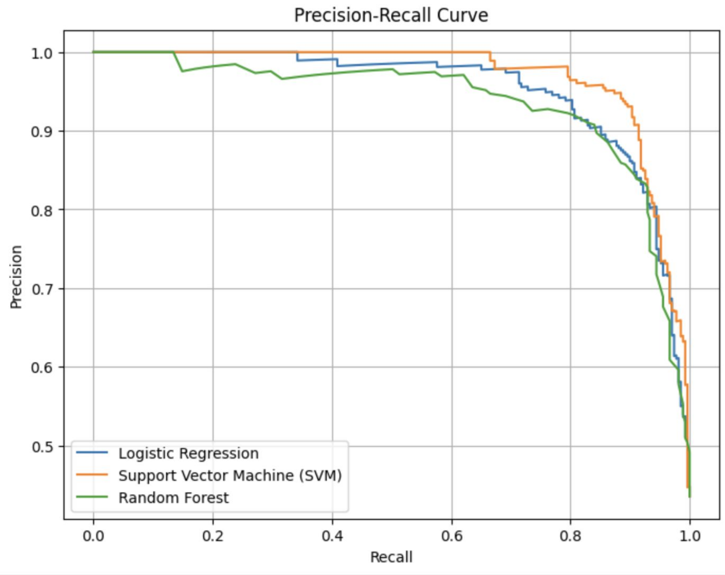
Results and Visualizations (Supervised Modeling)

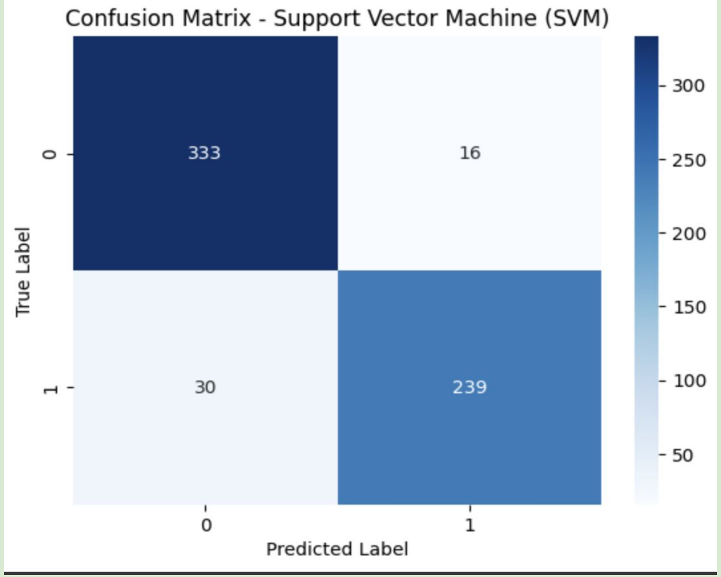
Supervised Approach was based of by using three traditional models, **Logistic regression, Random Forest** and **Support Vector Machine (SVM)**

Same stratified data was loaded to preserve the class balance

No Hyperparameter tuning was performed

SVM emerged as the best performer achieving high accuracy and strong separation between classes





Conclusion

Performance Summary of Supervised and Unsupervised Models. Below table shows Precision, Recall and F1 score for Real (Left column -Orange) and AI Gen (Right column- Green) Images

Model	Precision		Recall		F1 Score	
Logistic Regression	0.88	0.9	0.93	0.84	0.9	0.87
Support Vector Machine (SVM)	0.91	0.93	0.95	0.88	0.93	0.91
Random Forest	0.87	0.91	0.93	0.82	0.9	0.86
KMeans Majority Voting	0.68	0.61	0.72	0.56	0.7	0.58

While unsupervised methods reveal useful structure, they fall short in prediction performance without labeled guidance

Amongst all supervised models — especially **SVM with CLIP embeddings** — are highly effective for detecting AI-generated images. The high ROC AUC and F1-scores demonstrate **strong generalization and decision boundary clarity**, making this model suitable for real-world deployment