ML Assignment-Week 1 March 05, 2025

Machine Learning Model Performance Report

SHAILJA PATIL
24CS60R49

1. K-Nearest Neighbors (KNN)

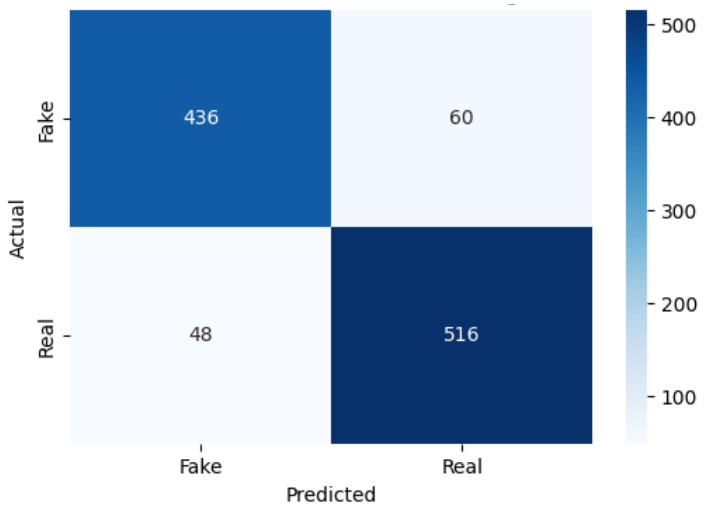
Best Hyperparameters:

- Algorithm: auto
- Metric: cosine
- Number of Neighbors: 15
- Weights: distance
- Classification Report on Test Set:

	Precision	Recall	F1-Score	Support
Fake	0.90	0.88	0.89	496
Real	0.90	0.91	0.91	564
Accuracy			0.90	1060
Macro Avg	0.90	0.90	0.90	1060
Weighted Avg	0.90	0.90	0.90	1060

Test Accuracy: 0.8981

CONFUSION MATRIX



2. Logistic Regression

Best Hyperparameters:

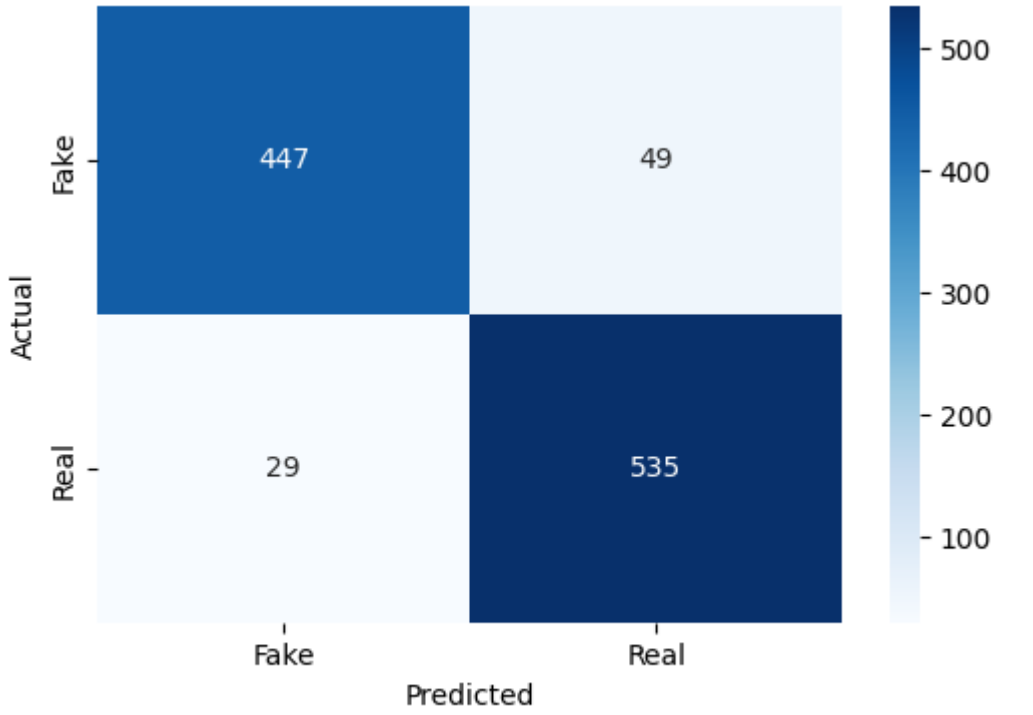
Default Parameters (Optimized via GridSearchCV)

Classification Report on Test Set:

	Precision	Recall	F1-Score	Support
Fake	0.94	0.90	0.92	496
Real	0.92	0.95	0.93	564
Accuracy			0.93	1060
Macro Avg	0.93	0.92	0.93	1060
Weighted Avg	0.93	0.93	0.93	1060

Test Accuracy: 0.9264

CONFUSION MATRIX



3. K-Means Clustering

Best K: 6 (Optimal based on Silhouette Score: 0.0119)

Silhouette Scores for Different K Values:

- K=2, Score=0.0064
- K=3, Score=0.0079
- K=4, Score=0.0098
- K=5, Score=0.0102

- K=6, Score=0.0119

Cluster Assignments: [0 0 2 4 4 4 3 3 2 0 4 4 4 4 4 3 3 4 4 3]

4. Support Vector Machine (SVM)

Best Hyperparameters:

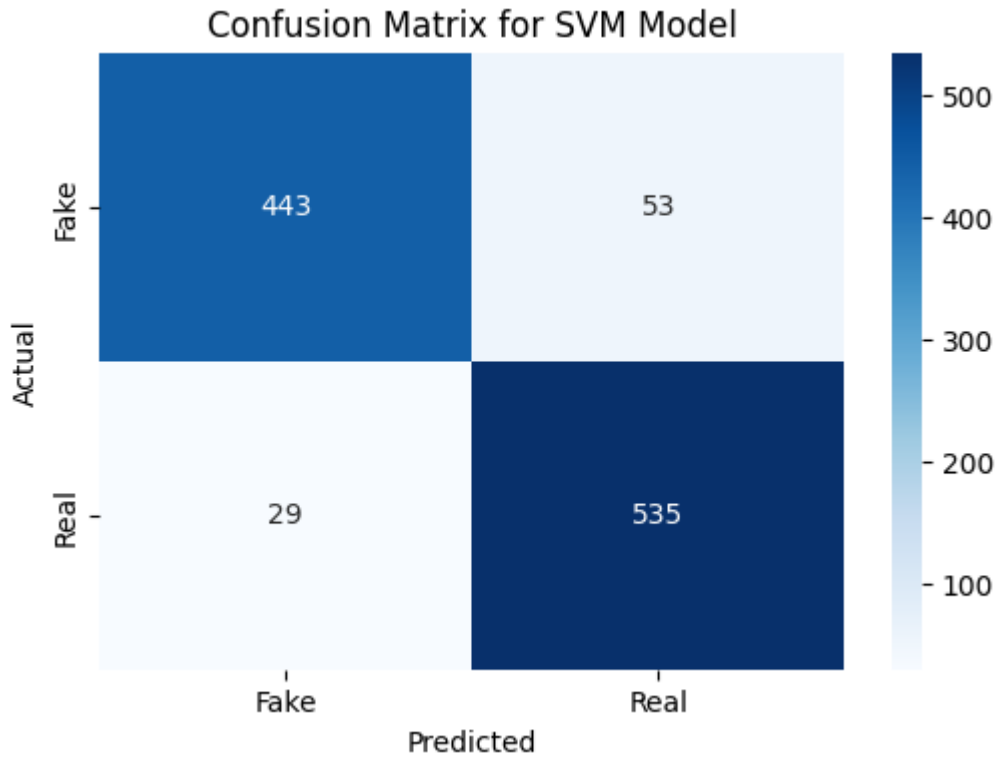
Default Parameters (Tuned via GridSearchCV)

Classification Report on Test Set:

	Precision	Recall	F1-Score	Support
Fake	0.94	0.89	0.92	496
Real	0.91	0.95	0.93	564
Accuracy			0.92	1060
Macro Avg	0.92	0.92	0.92	1060
Weighted Avg	0.92	0.92	0.92	1060

Test Accuracy: 0.9226

CONFUSION MATRIX



5. Neural Network (MLP)

Best Hyperparameters:

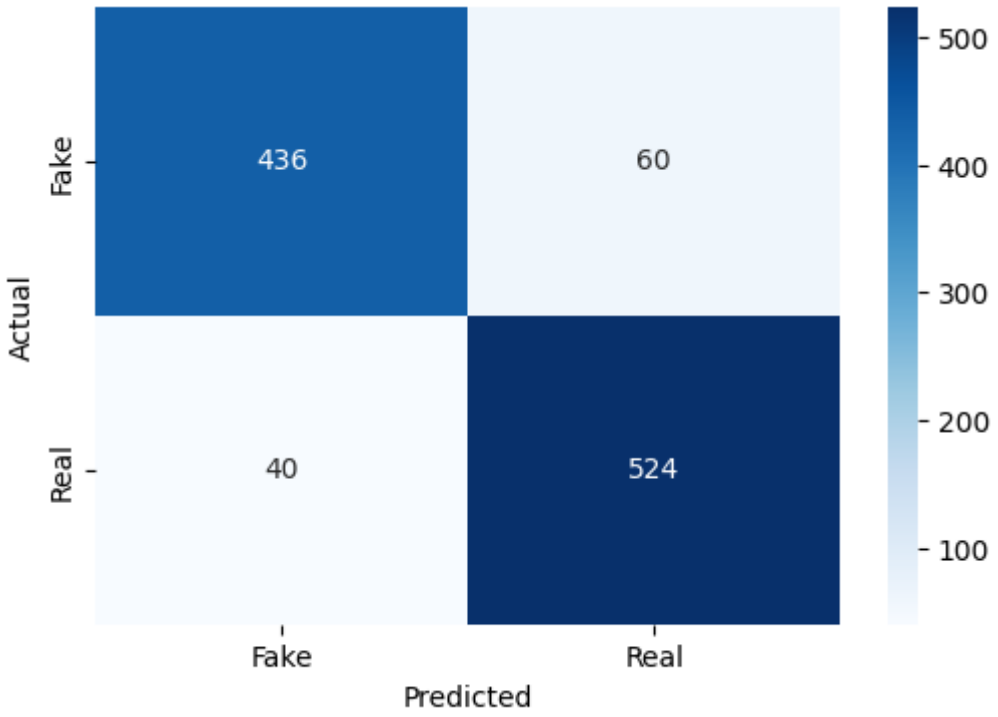
Optimized via RandomizedSearchCV

Classification Report on Test Set:

	Precision	Recall	F1-Score	Support
Fake	0.92	0.88	0.90	496
Real	0.90	0.93	0.91	564
Accuracy			0.91	1060
Macro Avg	0.91	0.90	0.91	1060
Weighted Avg	0.91	0.91	0.91	1060

Test Accuracy: 0.9057

CONFUSION MATRIX



6. Ensemble Learning - Gradient Boosting

Best Hyperparameters:

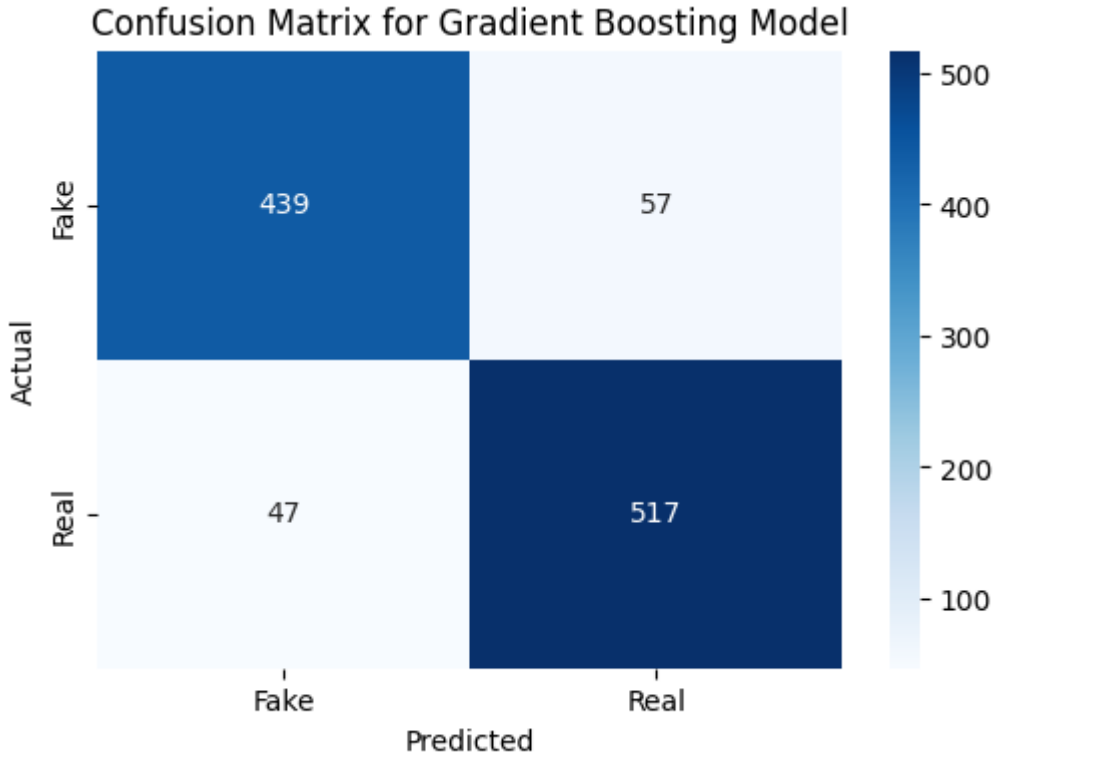
Optimized via GridSearchCV

Classification Report on Test Set:

	Precision	Recall	F1-Score	Support
Fake	0.90	0.89	0.89	496
Real	0.90	0.92	0.91	564
Accuracy			0.90	1060
Macro Avg	0.90	0.90	0.90	1060
Weighted Avg	0.90	0.90	0.90	1060

Test Accuracy: 0.9019

CONFUSION MATRIX



Conclusion

- Best Performing Model: Logistic Regression (Accuracy: 0.9264)
- Other High Performing Models:
  - SVM (Accuracy: 0.9226)
  - Neural Network (Accuracy: 0.9057)
  - Gradient Boosting (Accuracy: 0.9019)
- K-Means Clustering: Best K=6, but low Silhouette Score, indicating weak clustering performance for this task.
- KNN: Shows good accuracy (0.8981), but not as high as other models.
- Overall, Logistic Regression and SVM performed the best in classifying social media posts as real or fake, based on precision, recall, and F1-score.