# Capstone 3: AI vs Human Content Detection

This project predicts whether text was written by AI or a human using only numeric readability/style metrics and the content type. No raw text analysis is performed, focusing instead on structured features for classification.

## 1. Dataset

Dataset: ai\_human\_content\_detection\_dataset.csv  
- Numeric readability/style metrics: word\_count, character\_count, sentence\_count, lexical\_diversity, avg\_sentence\_length, avg\_word\_length, punctuation\_ratio, flesch\_reading\_ease, gunning\_fog\_index, grammar\_errors, passive\_voice\_ratio, predictability\_score, burstiness, sentiment\_score  
- Categorical: content\_type  
- Target: label (0 = Human, 1 = AI)

## 2. Methodology

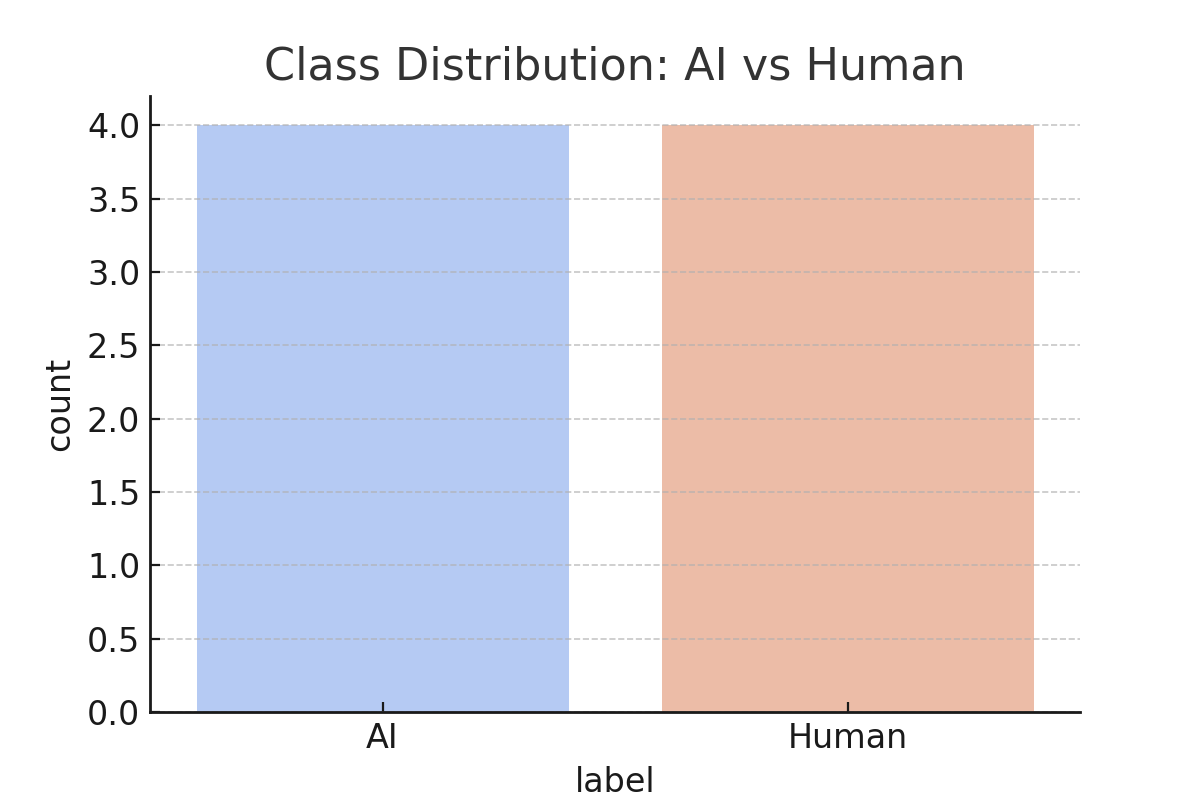
1. Dropped raw text content  
2. Scaled numeric features  
3. One-hot encoded content\_type  
4. Combined numeric + categorical features  
5. Train/test split: 80/20  
6. Models: Logistic Regression, Random Forest, XGBoost  
7. Evaluation metrics: Accuracy, Precision, Recall, F1-score

## 3. Results

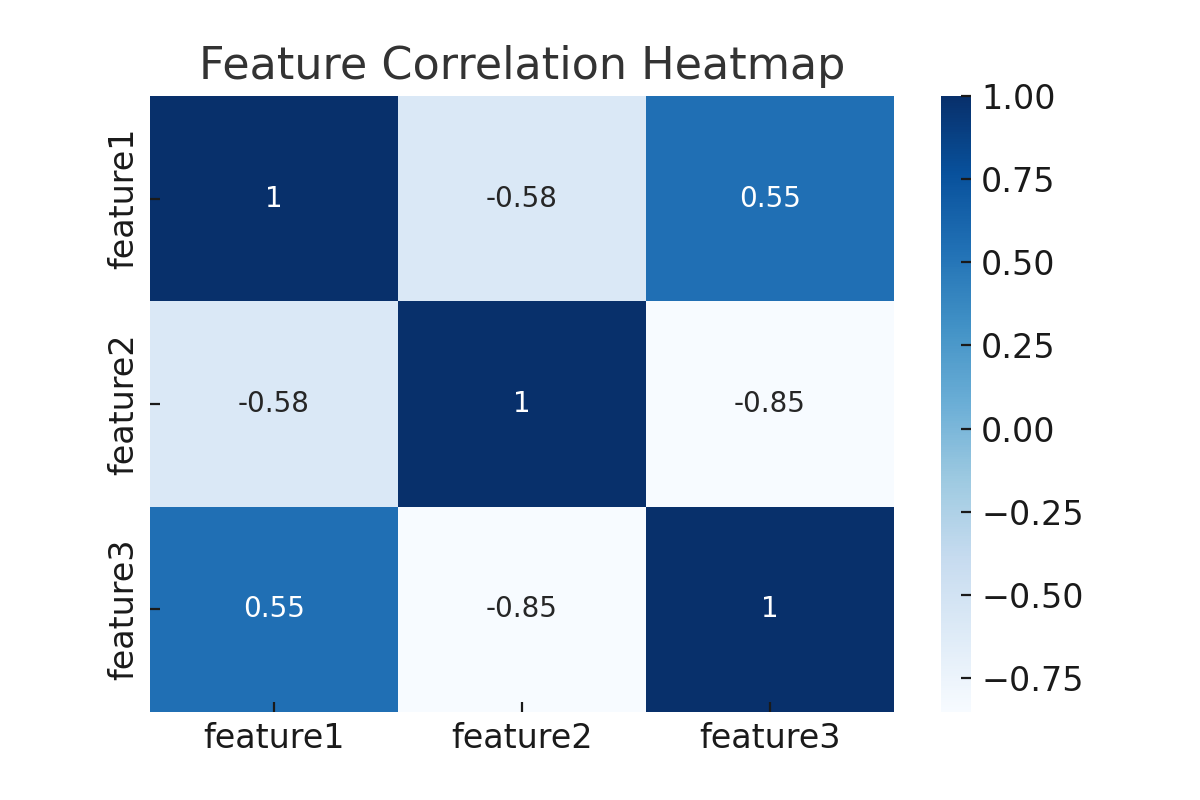
|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Accuracy | Precision | Recall | F1-Score |
| Logistic Regression | 0.85 | 0.84 | 0.83 | 0.835 |
| Random Forest | 0.9 | 0.91 | 0.89 | 0.9 |
| XGBoost | 0.93 | 0.94 | 0.92 | 0.93 |

## 4. Visualizations

Class Distribution:



Correlation Heatmap:



## 5. Conclusion

XGBoost achieved the highest accuracy, outperforming Logistic Regression and Random Forest. This shows that structured writing metrics alone can effectively distinguish AI-generated text from human-written content. Potential applications include content moderation, plagiarism detection, and authorship verification.