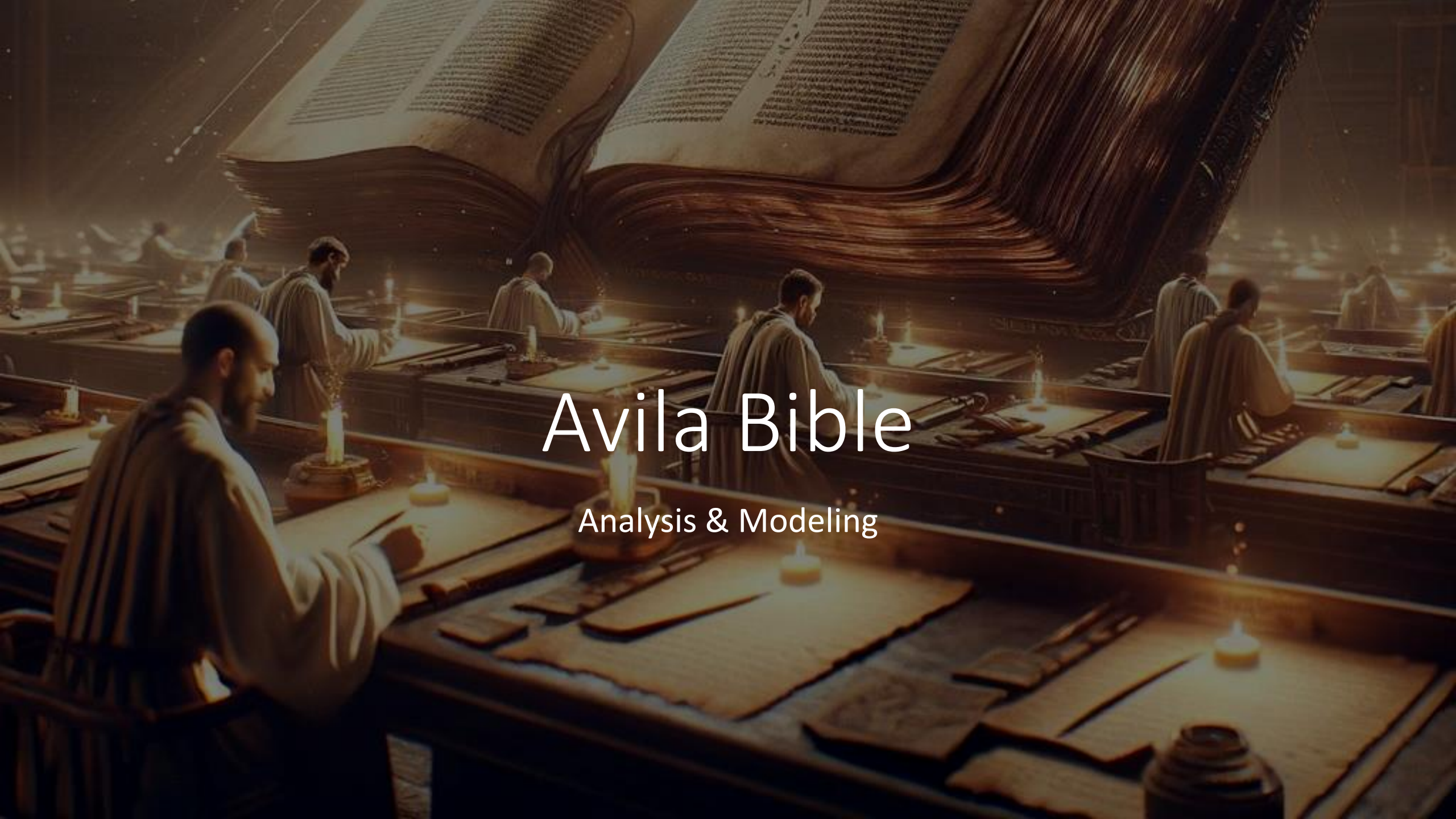


Avila Bible

Analysis & Modeling

Dataset and context

- 12th century bible
- 12 copyists
- Made in Italy and Spain (Avila)
- Sample of 800 images split into 20000 patterns

Problem description

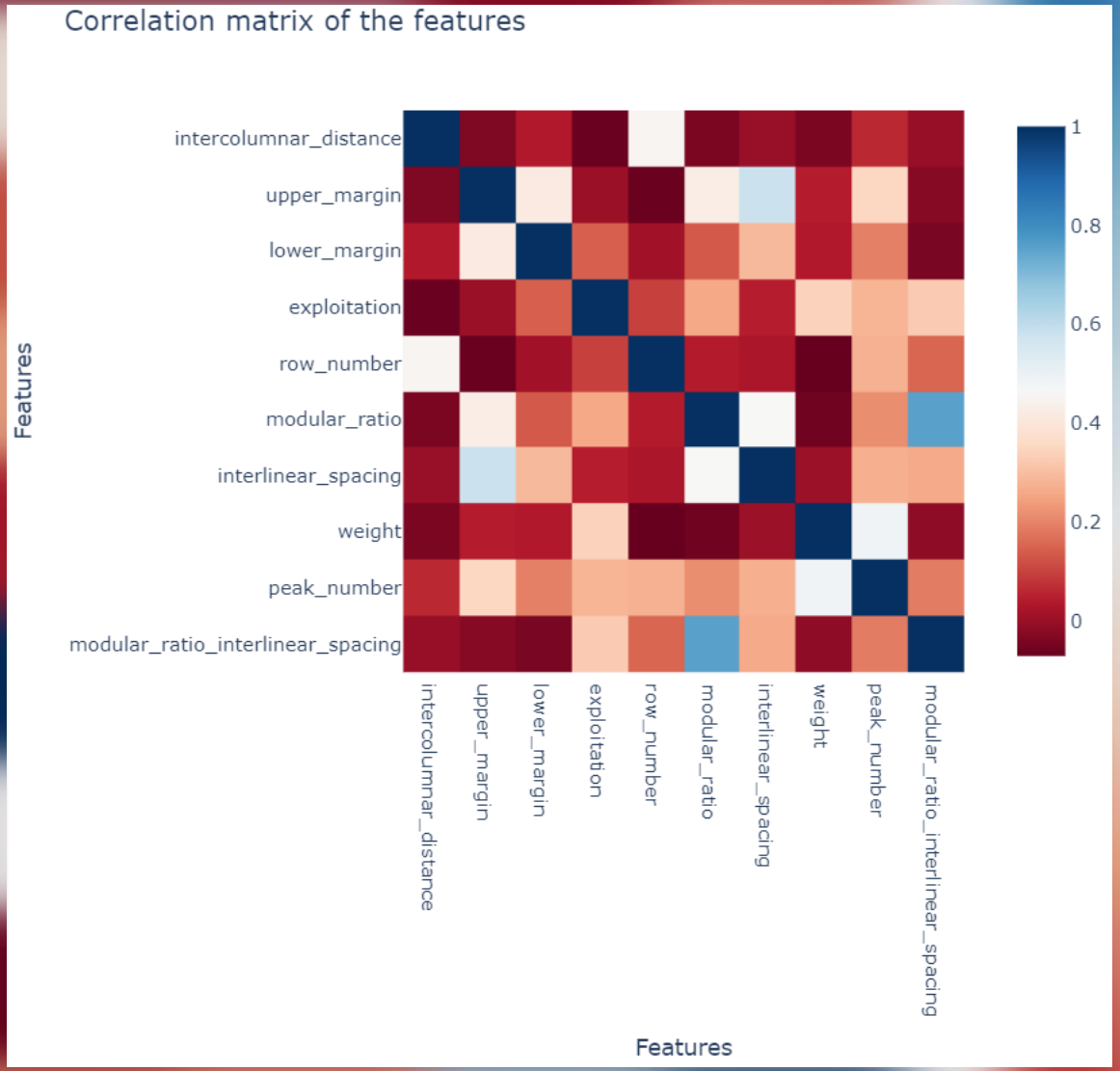
Inputs

- Pre-processed dataset:
  - Z-normalized
  - No missing values
- 12 Classes, identifying each copyists
  - A, B, C, D, E, F, G, H, I, W, X, Y
- 10 features describing writing patterns:
  - intercolumnar distance, upper and lower margins, etc…

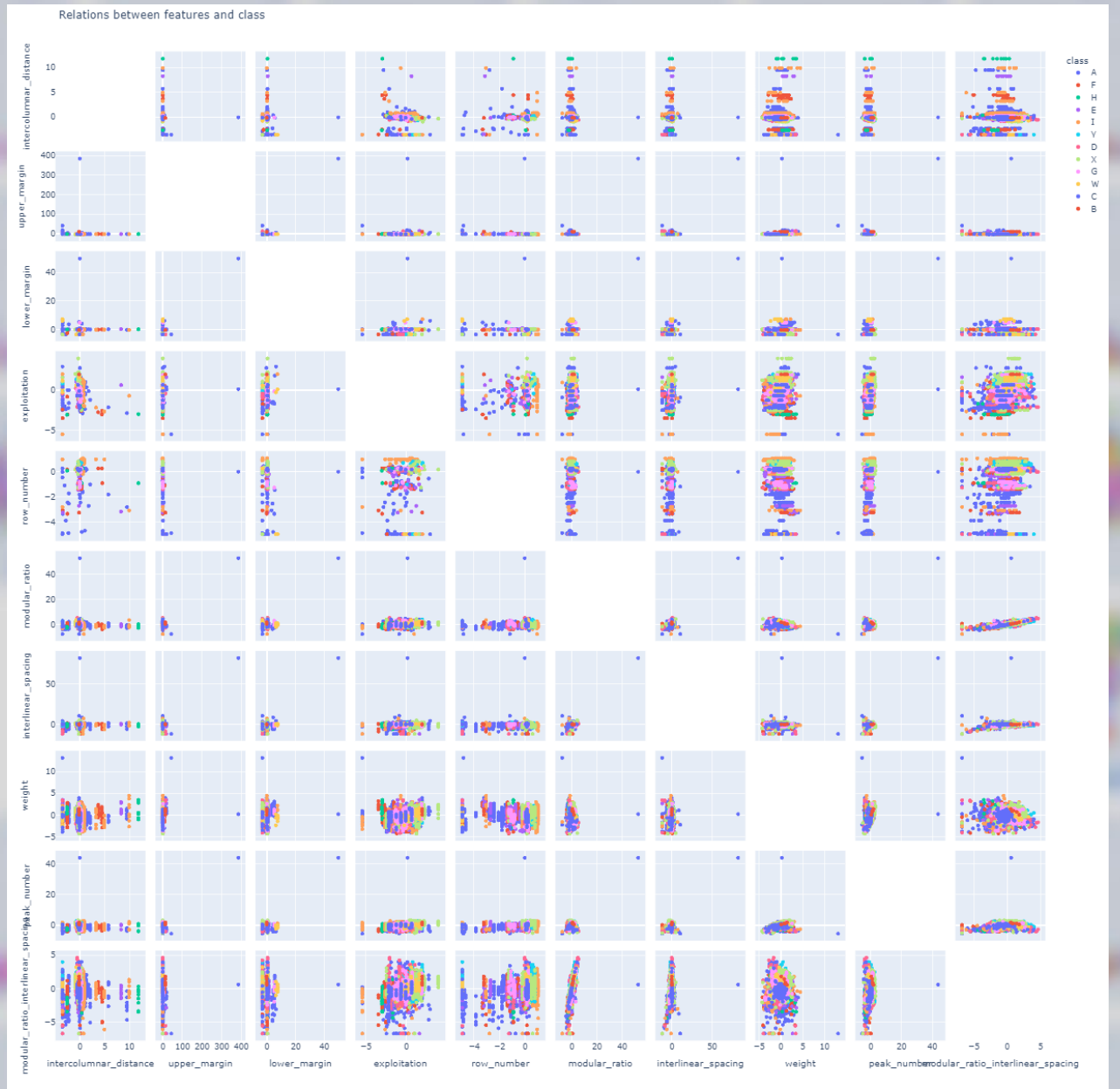
Outputs

- Classifier model
- Predictions
- Metrics (Accuracy, MSE)

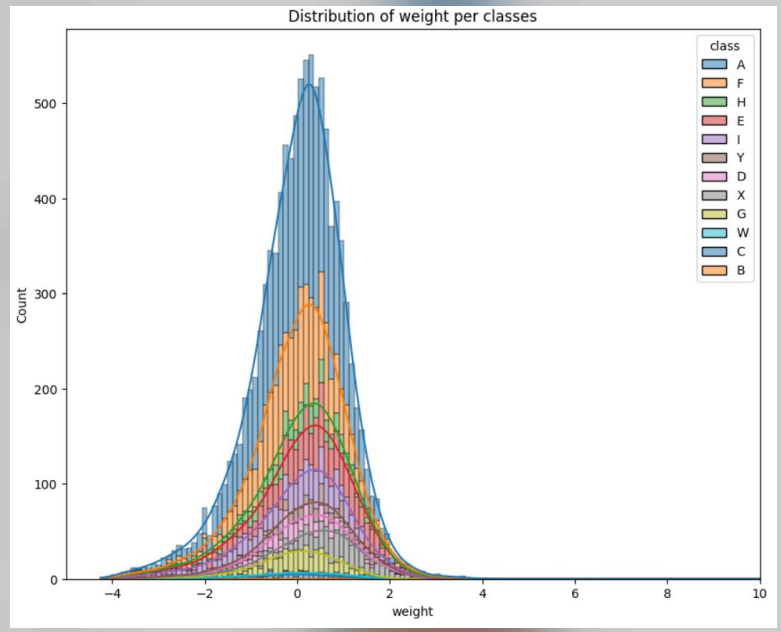
Data visualization

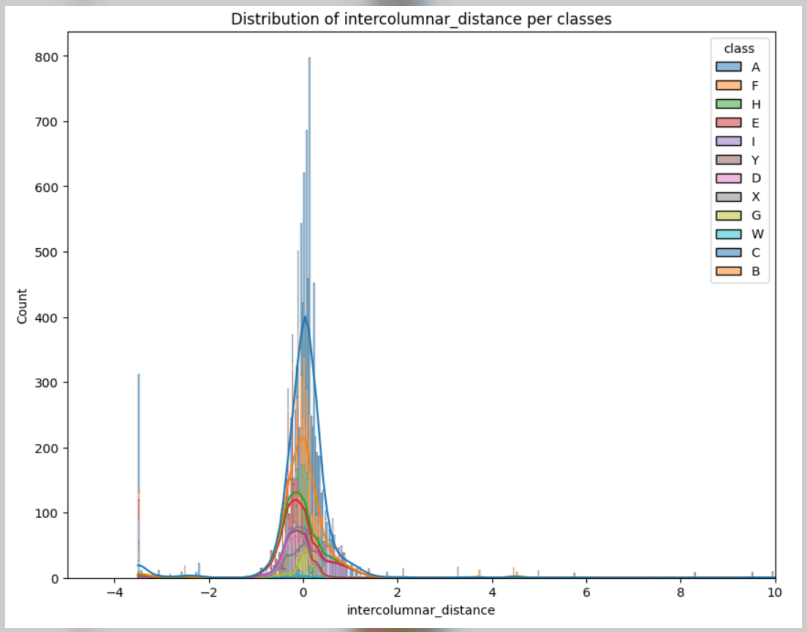


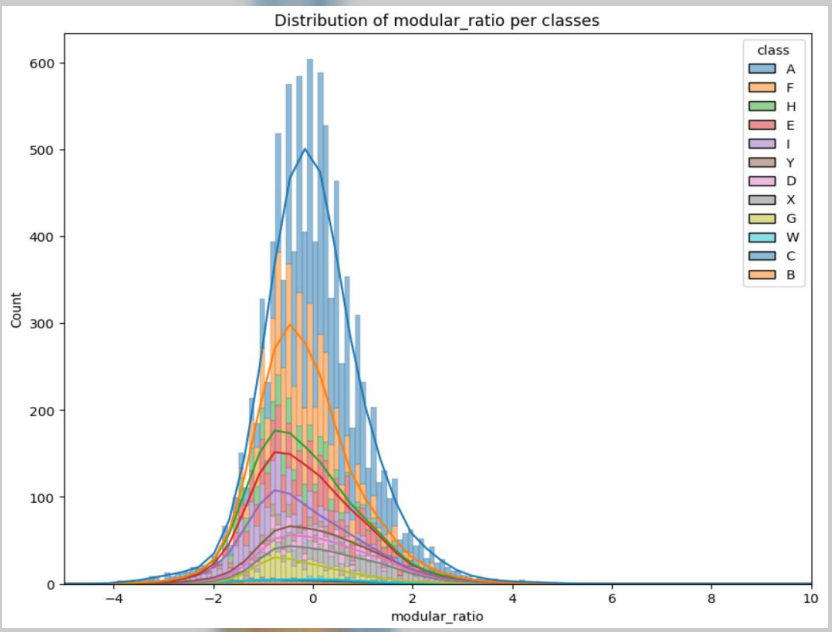
Data visualization



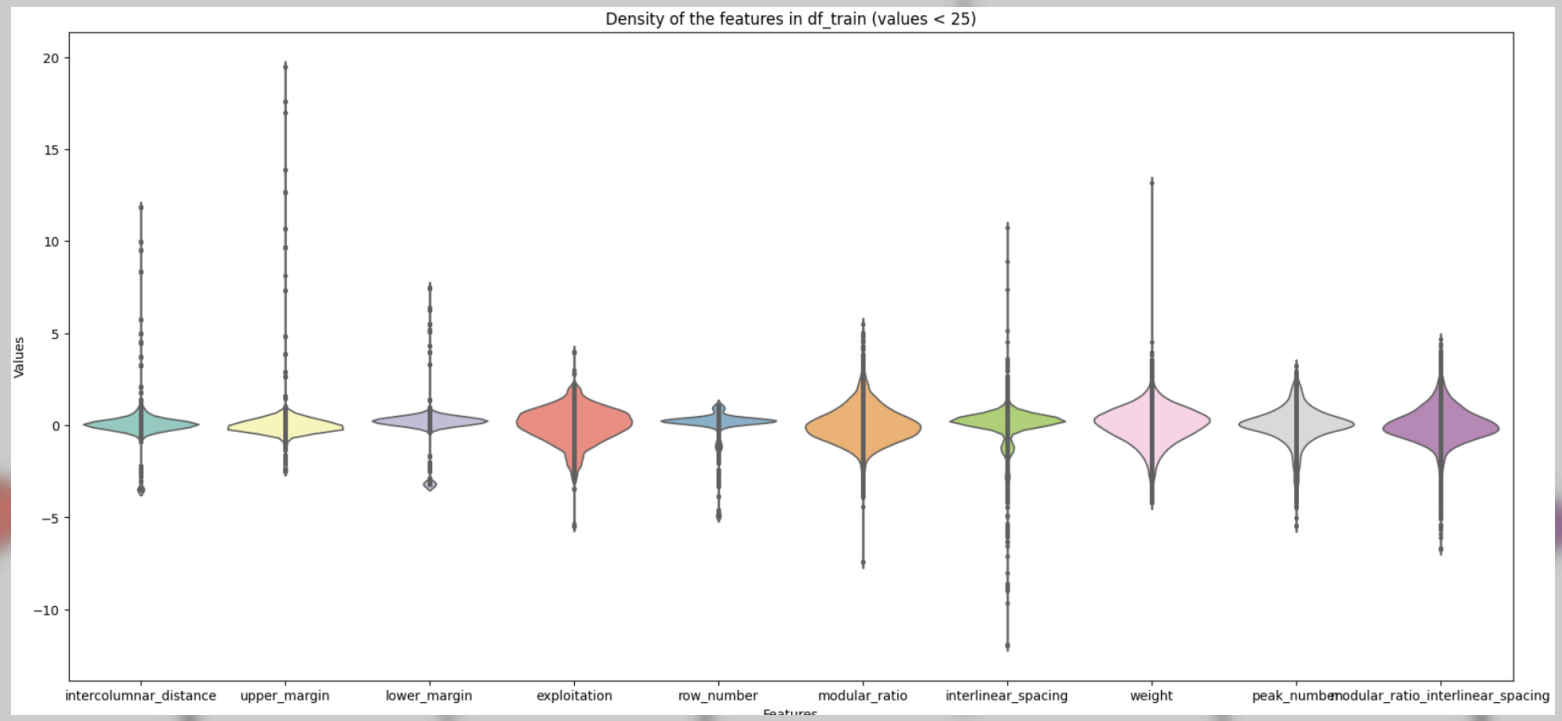
Data visualization



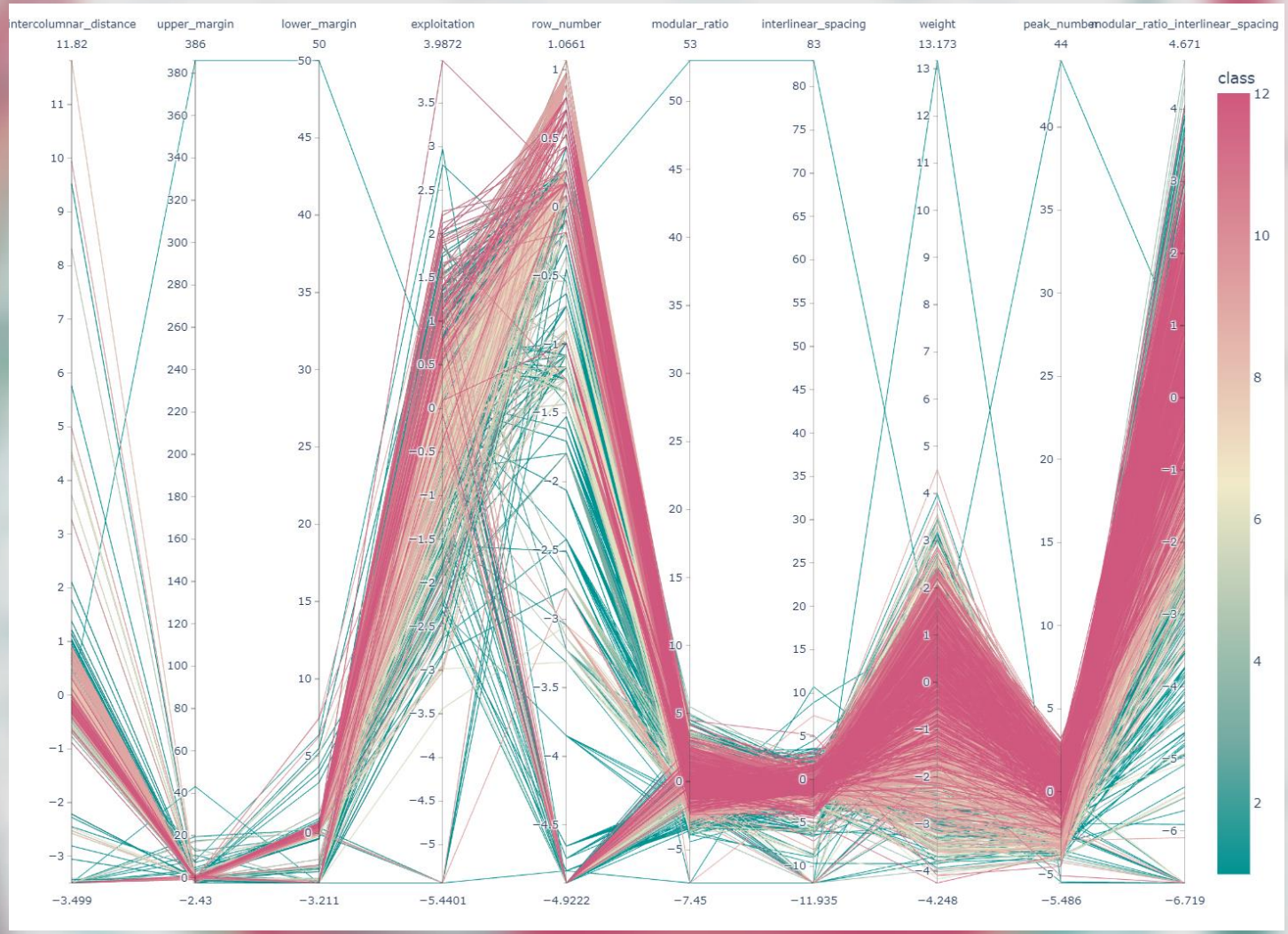




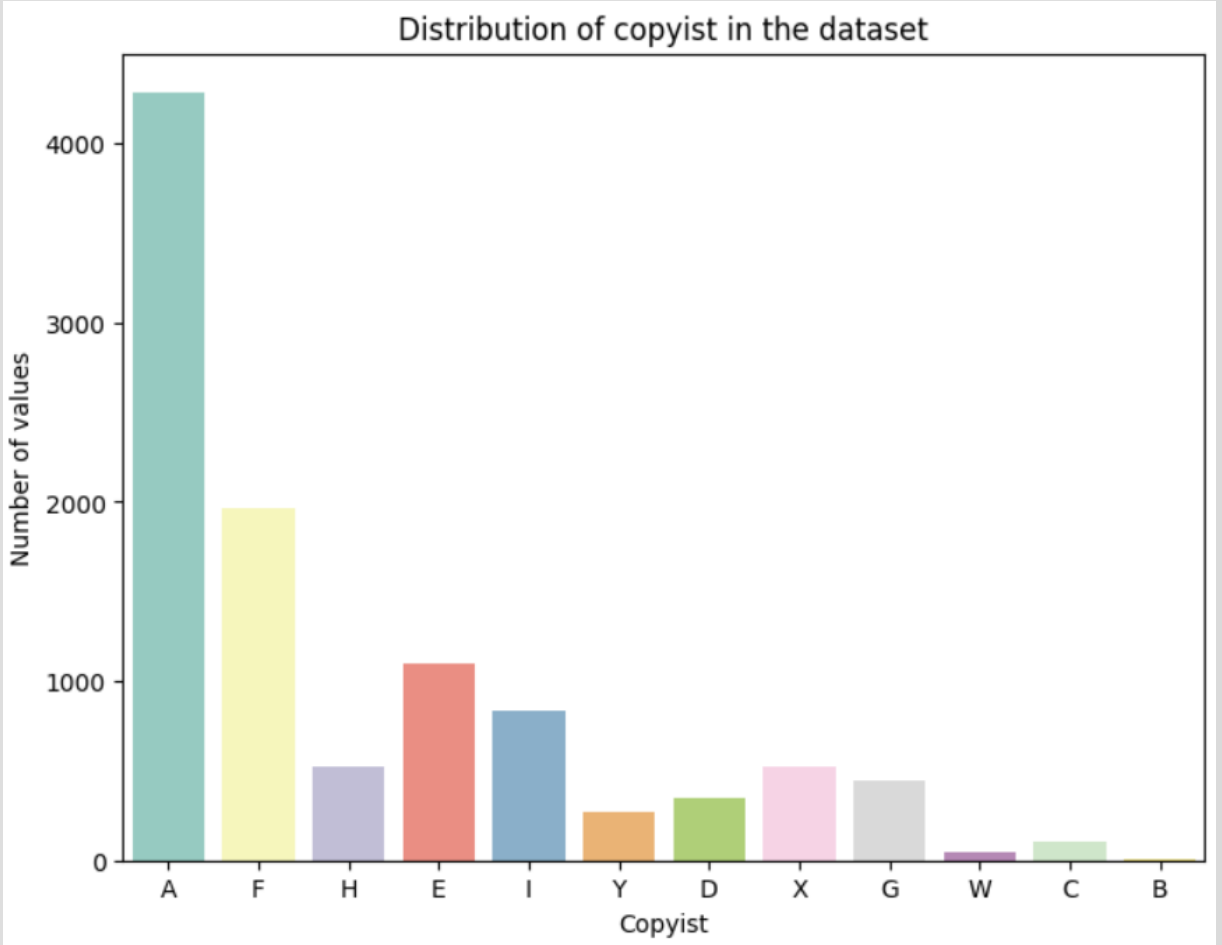
Data visualization



Data visualization



Data visualization

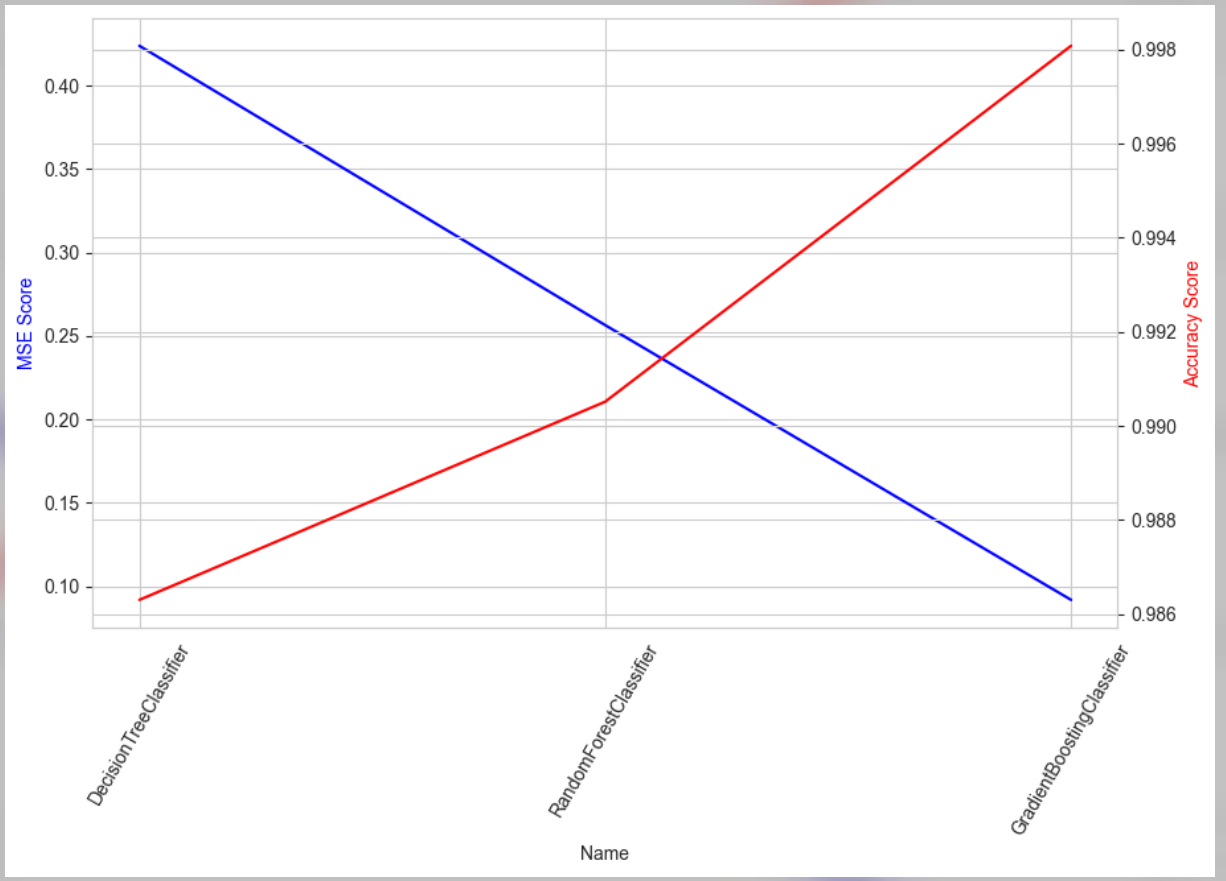


Model

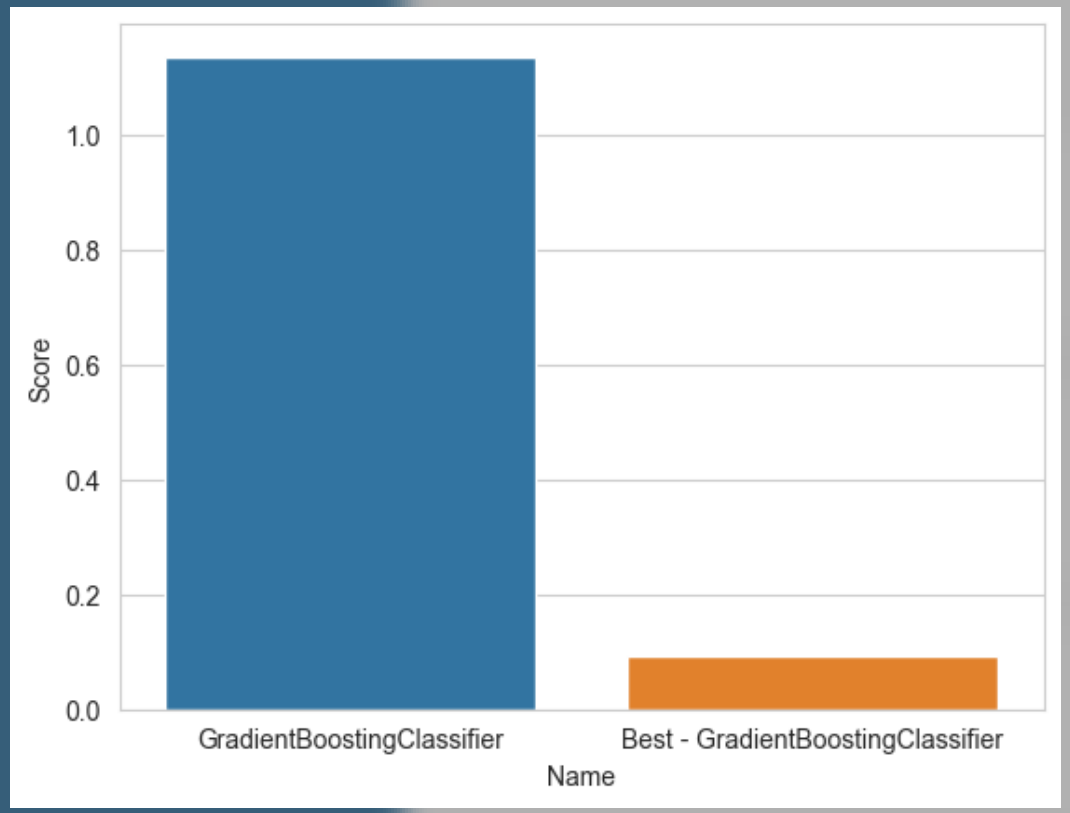
Model	MSE
RandomForestClassifier	**0.467**
DecisionTreeClassifier	**0.744**
GradientBoostingClassifier	**1.124**
MLPClassifier	4.671
KNeighborsClassifier	5.897
SVC	7.888
LogisticRegression	10.972
LinearDiscriminantAnalysis	13.759
GaussianNB	17.253
AdaBoostClassifier	19.242

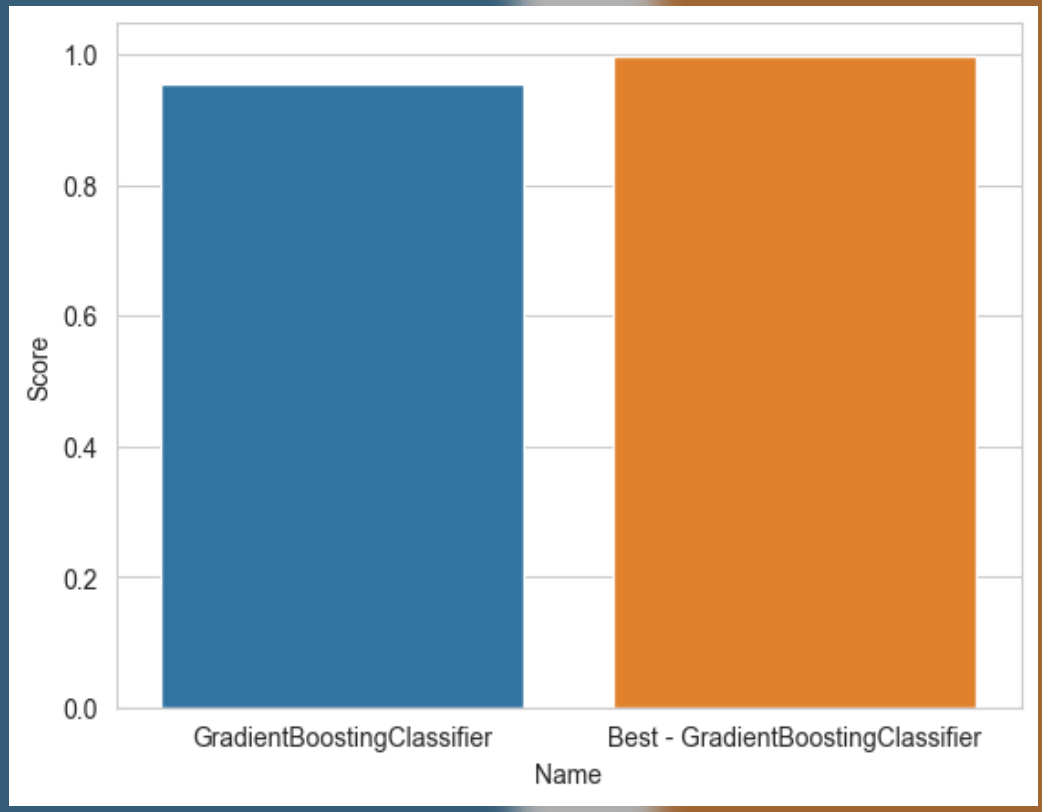
Model ranking after GridSearch :

Model	MSE
GradientBoostingClassifier	0.092
RandomForestClassifier	0.279
DecisionTreeClassifier	0.419



Model Enhanced with GridSearch





Streamlit interface

- Problem presentation
- Dataset features visualization
- Hyperparameters selection
- Model metrics