1. **Data Loading and Sampling**:
   1. The HIGGS dataset is loaded, and a 1% sample is taken to manage computational load and focus on model performance over accuracy.
2. **Data Preprocessing and Feature Engineering**:
   1. The dataset is standardized using StandardScaler to normalize feature values.
   2. Polynomial features are generated up to degree 2 (without bias) to capture interactions between features, increasing the feature set from 28 to 406 dimensions.
   3. SelectKBest with the ANOVA F-value is used to reduce dimensionality, selecting the 10 most relevant features.
3. **Data Splitting**:
   1. The preprocessed data is split into training and testing sets (70% training, 30% testing) for evaluation.
4. **Model Training and Cross-Validation**:
   1. A Linear SVM model is trained on the reduced dataset.
   2. Cross-validation (5-fold) is used to evaluate the model on training data, capturing metrics such as accuracy, precision, recall, F1 score, and ROC AUC.
5. **Model Evaluation on Test Data**:
   1. The model’s performance is assessed on the test set, reporting accuracy, precision, recall, F1 score, and AUC.
   2. Additionally, training and prediction times are measured for time complexity analysis.

**Results Summary:**

* **Kernel Methods Used**:
  + **Linear SVM:**

Original number of features: (110000, 28)

New number of features with polynomial features: (110000, 406)

Cross-Validation Results (Linear SVM):

Accuracy: 0.6351038961038962

Precision: 0.6182215449337651

Recall: 0.8035944788491746

F1 Score: 0.6988199165461573

Area\_ROC: 0.6736944022470748

Test Set Evaluation (Linear SVM):

Accuracy: 0.6358484848484849

Precision: 0.6193308550185873

Recall: 0.8104040288428522

F1 Score: 0.702099705000124

AUC: 0.6248980082382495

Linear SVM Training Time: 285.93 seconds

Linear SVM Prediction Time: 29.41 seconds

* Other potential kernels to consider for comparison:
  + **Radial Basis Function (RBF)**: Known for handling non-linear relationships effectively.
  + **Polynomial Kernel**: Suitable for capturing interactions up to a specified degree.
  + **Sigmoid Kernel**: Often used for binary classification with data that is not linearly separable.
* **Hyperparameter Variations**:
  + **C (Regularization Parameter)**: Controls the trade-off between achieving a high margin and correctly classifying training examples.
  + **Gamma (for RBF and Polynomial Kernels)**: Influences the decision boundary’s reach and flexibility.
  + **Degree (for Polynomial Kernel)**: Specifies the polynomial's degree to capture feature interactions.
* **Stochastic Gradient Descent SVM:**

Cross-Validation Results (SGDClassifier):

Accuracy: 0.6305454545454545

Precision: 0.6237754671295578

Recall: 0.7544368457033552

F1 Score: 0.6824685529530137

Area under ROC: 0.6677014500057961

Test Set Evaluation (SGDClassifier):

Accuracy: 0.6339393939393939

Precision: 0.6136823469903895

Recall: 0.8331807256495365

F1 Score: 0.706781882615661

AUC: 0.6214402919758697

Best Parameters for SGD SVM: {'alpha': 0.001, 'max\_iter': 2000}

Best Cross-Validation Score: 0.6342597625687699

SGD SVM Training Time: 0.33 seconds

SGD SVM Prediction Time: 0.00 seconds

* **Polynomial SVM:**

Polynomial SVM (degree=2) Accuracy: 0.6053939393939394

Polynomial SVM (degree=3) Accuracy: 0.6404242424242425

Polynomial SVM (degree=4) Accuracy: 0.6296969696969696