**Project Title: Bitcoin Market Sentiment project**

**Objective:**

The aim of this project is to predict the **sentiment of cryptocurrency trades** ('Extreme Fear', 'Extreme Greed', 'Fear', 'Greed', 'Neutral') based on transactional and market features,helping analysts and traders understand market behavior and decision patterns.

**Dataset Overview:**

* **Total Records:** 211,224
* **Features:** 17 columns, including account details, coin, execution price, size, side, direction, fees, and FearGreedValue.
* **Target Variable:** classification (Positive, Negative, Neutral)
* **Preprocessing:**
  + Dropped unnecessary timestamp columns (SentimentDate, SentimentTimestamp, Timestamp).
  + Checked for missing values and handled them. Only FearGreedValue and classification had minor missing values (6 records).
  + Label encoding applied for categorical columns (classification, Direction, Side, Coin).
  + Features scaled using **Robust Scaler** and transformed with **PowerTransformer** for better model performance.

**Model Selection:**

* **Models Evaluated:** Logistic Regression, Random Forest, Gradient Boosting, K-Nearest Neighbors, Gaussian Naive Bayes, XGBoost, Decision Tree.
* **Cross-validation (5-fold)** performed for each model to compare accuracy, standard deviation, and ROC-AUC scores.

**Cross-validation Summary:**

| **Model** | **CV Accuracy** | **Std Deviation** | **ROC-AUC** | **Rank** |
| --- | --- | --- | --- | --- |
| RandomForestClassifier | 1.000 | 0.000 | 1.000 | 1 |
| GradientBoostingClassifier | 1.000 | 0.000 | 1.000 | 1 |
| XGBClassifier | 1.000 | 0.000 | 1.000 | 1 |
| DecisionTreeClassifier | 1.000 | 0.000 | 1.000 | 1 |
| LogisticRegression | 0.9875 | 0.0010 | 0.9999 | 5 |
| KNeighborsClassifier | 0.9338 | 0.0125 | 0.9897 | 6 |
| GaussianNB | 0.8817 | 0.0019 | 0.9895 | 7 |

**Model Chosen:** **Logistic Regression**

**Reason for Choice:**

1. High accuracy (0.987) on balanced and preprocessed dataset.
2. Robust to overfitting compared to tree-based models which gave 100% accuracy (possible overfitting).
3. Efficient and faster to train on large datasets (~2 lakh records).
4. Easy to interpret coefficients for feature importance and insights.

**Final Model Performance:**

**Logistic Regression on Test Set:**

* **Accuracy:** 0.9857
* **Confusion Matrix:**

[[ 4277 0 7 0 0]

[ 0 7874 0 0 0]

[ 237 0 12163 0 47]

[ 0 3 0 9772 277]

[ 0 0 34 0 7553]]

* **Classification Report:**

precision recall f1-score support

0 0.95 1.00 0.97 4284

1 1.00 1.00 1.00 7874

2 1.00 0.98 0.99 12447

3 1.00 0.97 0.99 10052

4 0.96 1.00 0.98 7587

accuracy 0.99 42244

macro avg 0.98 0.99 0.98 42244

weighted avg 0.99 0.99 0.99 42244

**Observations:**

* Model predicts all classes with high precision and recall.
* Slight misclassification in class 2 and 3, but overall performance is excellent.
* No major overfitting observed due to cross-validation and balanced dataset.

**Model Testing & Validation:**

* Compared y\_test vs y\_pred using **confusion matrix** and **classification report**.
* Verified predictions by decoding numerical labels to original sentiments ('Extreme Fear', 'Extreme Greed', 'Fear', 'Greed', 'Neutral') using **LabelEncoder**.
* Ensured the model generalizes well on unseen test data.

**Model Deployment:**

* Model saved using **Pickle** for future predictions:
* Ready for integration into trading analysis systems for real-time sentiment prediction.

**Conclusion:**

The **Logistic Regression model** provides an accurate, interpretable, and robust solution for predicting sentiment in cryptocurrency trading data. The preprocessing, feature scaling, and careful model selection ensure high reliability for practical use.