1. **Imbalanced Binary Classification:** A classification problem where one class (the majority class) has significantly more instances than the other class (the minority class). In the context of the document, it refers to credit card fraud detection, where fraudulent transactions are much rarer than legitimate ones.
2. **Precision-Recall Tradeoff:** The inverse relationship between precision and recall. Increasing precision (reducing false positives) often leads to decreased recall (increasing false negatives), and vice versa. The optimal balance depends on the specific application.
3. **RobustScaler:** A data preprocessing technique that scales features to a common range, making them less sensitive to outliers. It's used to normalize the 'Amount' and 'Time' features in the credit card dataset.
4. **Confusion Matrix:** A table used to evaluate the performance of a classification model by comparing actual and predicted class labels.
5. **Accuracy:** The proportion of correctly classified instances out of the total instances. It's not always the best metric for imbalanced datasets.
6. **Precision:** The proportion of true positive predictions out of all positive predictions. It measures how often the model is correct when it predicts the positive class (fraudulent in this case).
7. **Recall:** The proportion of true positive predictions out of all actual positive instances. It measures how well the model identifies all actual positive instances. Also known as sensitivity.
8. **F1 Score:** The harmonic mean of precision and recall. It provides a single metric that balances both precision and recall.
9. **Type I Error:** Incorrectly rejecting the null hypothesis when it is actually true. Equivalent to a false positive.
10. **Type II Error:** Incorrectly failing to reject the null hypothesis when it is actually false. Equivalent to a false negative.
11. **class\_weight='balanced':** A parameter in machine learning models that adjusts the weights of different classes to address class imbalance. It gives more weight to the minority class during training.
12. **solver='liblinear':** An algorithm used for optimization in machine learning models. It's suitable for smaller datasets and binary classification problems.