

SMOTE (Synthetic Minority Over-sampling Technique) is a technique used in machine learning to address class imbalance in datasets. When you have a dataset where one class is significantly underrepresented compared to another, SMOTE helps by generating synthetic examples of the minority class to balance the class distribution.

Here's a high-level overview of how SMOTE works:

1. **Identify Minority Class Samples:** SMOTE starts by identifying samples of the minority class in the dataset.
2. **Generate Synthetic Samples:** For each minority class sample, SMOTE generates synthetic samples by creating new examples that are combinations of the original sample and its nearest neighbors in the feature space.
3. **Add Synthetic Samples:** These synthetic samples are then added to the dataset, increasing the number of minority class examples and helping the model to learn more about the minority class.

SMOTE can be particularly useful for improving the performance of classifiers when dealing with imbalanced datasets, making it a popular technique in data preprocessing for machine learning tasks.