Link to source: <https://www.sciencedirect.com/science/article/pii/S0020025524014841>

Project Idea: Going to modify SMOTE by combining class density feature analysis.

Synthetic Minority Over-sampling Technique (SMOTE) is a popular algorithm used to address this issue by generating synthetic samples for the minority class. However, SMOTE does not consider the density distribution of classes, which can lead to two significant problems:

1. Synthetic Data Overlap: SMOTE may generate synthetic samples in regions where the majority class dominates, leading to class overlap and reducing the model's discriminative ability.
2. Ignoring Class Density Variation: Real-world datasets often exhibit variation in the density of minority class samples. Some regions may be sparse while others are dense, making these easier to learn.. SMOTE treats all regions equally, which can result in suboptimal synthetic sample generation.

By analyzing class density features, the algorithm can avoid generating synthetic samples in regions dominated by the majority class, thus reducing the risk of class overlap and enhancing model performance. To implement this, we plan to follow the research paper linked above, which provided multiple different algorithms for possible implementation.

Wait what about this:

<https://machinelearningmastery.com/cost-sensitive-learning-for-imbalanced-classification/>

<https://juandelacalle.medium.com/i-declare-myself-the-1-enemy-of-over-undersampling-smote-and-adasyn-heres-why-how-i-5889b5073419>