Experimental Settings

A. **Datasets**

In our study, we evaluate the proposed method with four UCI ML and one Kaggle reciprocity datasets. The datasets are summarized in a table.

- **Online Shoppers Purchasing Intention Dataset(OS)** [Sakar and Kastro, 2018] sourced from UCI, it contains a mix of real and integer both type attributes which is related to user behaviour on an e-commerce website. It is designed for binary classification tasks including customer purchase and purchase prediction.

- **MAGIC Gamma Telescope Dataset (MA)** [Bock, 2007] from UCI, this dataset simulates gamma particle detection in a gamma telescope. The main purpose of this dataset is to distinguish actual gamma signals from background cosmic rays, helping in particle classification tasks in astrophysics.

- **Smoking and Drinking Dataset with Body Signal** (**SD**) [Her, 2023] sourced from the Korean National Health Insurance Service and hosted on Kaggle, includes physiological body signal features. Its main objective is to help classification tasks that distinguish between smokers and drinkers.

- **Car Evaluation Dataset(CE)** [Bohanec, 1997] from UCI, this dataset involves evaluation of cars based on various attributes like price, maintenance cost and safety. The main purpose of this dataset is to classify cars into classes such as unacceptable, acceptable, or excellent, supporting decision making in automotive sectors.

- **Statlog(German Credit Card) Dataset (ST)** [Hofmann, 1994] available on UCI, this dataset is a benchmark dataset which involves various attributes like past loan history, savings and employment details. Its main objectives are to predict loan defaults, supporting research in fraud detection, financial risk evaluation etc.

Datasets	*Instances*	*Features*	*Classes*
OS	12330	17	2
MA	19020	10	2
SD	991346	23	2
CE	1728	6	4
ST	1000	20	2

B. Downstream Tasks. We assess the proposed model on a diversity of classification tasks, utilizing algorithms such as Random Forests (RF), Logistic Regression (LR), Naive Bayes (NB), K-Nearest Neighbors (KNN), and Multilayer Perceptrons (MLP). The performance of each classifier is compared with and without the integration of our method to evaluate its effectiveness. To assess the effectiveness of our strategy, we evaluate the performance of each classifier with and without our strategy.

C. Baseline Models. To demonstrate the effectiveness of our proposed CAFWT technique, we compare it against four widely recognized baseline techniques. Specifically, LASSO and TabTransformer are used for feature preprocessing, whereas Weighted Bootstrapping (WB) and Undersampling (USP) are used for sample weight preprocessing.

- *Undersampling (USP)* is a method that balances the dataset by reducing the number of samples from the majority class, aligning it with the size of the minority class. This technique contributes to mitigate bias and improves model fairness in model predictions. In our study, the undersampling ratio is based on the observed frequency of each category.

- *Least Absolute Shrinkage and Selection Operator (LASSO)* is a regression technique that improves model performance by absolute magnitude of the regression coefficients. This technique aids in filtering out irrelevant features, simplifying the model structure and enhances the prediction accuracy.

- *Weighted Bootstrapping (WB)* [Barbe and Bertail, 1995] is a resampling technique that assigns different weights in dataset instances, affecting their selection during the resampling process. This technique is particularly effective in operating imbalanced datasets with more importance to classes presented. In our setup, the weight assignment is directly impacted by class distribution.

- T*abTransformer (TabT)* [Huang et al., 2020] is a transformer-based solution that is designed for tabular datasets. It converts categorical features into embeddings and applies attention techniques to accept intricate feature relationships. In our experiments, TabTransforme we used TabTransformer to process categorical features, generating influential representations that improve model's learning capabilities.

D. Metrics. To assess the effectiveness of our proposed technic, we rely on the following core performance metrics:

⬨ **Overall Accuracy (Acc)**: It represents the ratio of correctly classified instances both true positive and true negative over the whole number of components in the dataset.
⬨ **Precision (Prec)**: It indicates the proportion of true positive predictions to all positive predictions made by the model. It shows how reliable the model is when it predicts positive class.
⬨ **Recall (Rec)**: Also referred to as sensitivity, this metric evaluate the model's ability to accurately identify all actual positive instances. It shows how to accurately evaluate the positive class.
⬨ **F-Measure (F1 Score)**: It gives a balanced average of accuracy and recall, gives the balanced evaluation of the model accuracy, making it especially effective when dealing with class imbalance.

Colab Link Of All Code