

# Assignment 5

Dataset to be used: <https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset/data>

- Create a function to apply 3-5 basic text cleaning methods and apply it on the text column.
- Create BOW features for the given text data.
- Train a Naive Bayes model + random forests + xgboost.
  - Train once with cleaning and once without cleaning.
- Repeat the same with Tf-IDF.
- Explore ensemble methods for classification.