# Title: Leveraging Machine Learning for Automated Consumer Complaint Classification

## Project Overview:

In the fast-paced world of customer service, timely and accurate responses to consumer complaints are critical for maintaining customer satisfaction. This project aims to automate the classification of consumer complaints into predefined categories, thus improving the efficiency of complaint resolution processes.

The project uses a variety of machine learning algorithms, including RandomForest, XGBoost, Bernoulli Naive Bayes, DecisionTree, CatBoost, RandomForestClassifier, XGBClassifier, and LightGBM, to categorize complaints based on their textual content. Furthermore, the project leverages the power of stacked generalization (ensemble learning) to improve the robustness and accuracy of the classification process.

## Workflow:

**1. Data Preparation:** This step involves loading the data, handling missing values, and carrying out preliminary data analysis.

**2. Text Preprocessing:** The project uses TF-IDF Vectorizer to convert the complaint text into a matrix of TF-IDF features.

**3. Training Various Models:** We train a variety of machine learning models, each of which approaches the classification problem in a unique way. This diversification increases the chance of finding a model that can accurately classify the consumer complaints.

**4. Model Evaluation:** Each model is evaluated based on its accuracy. This is done through a classification report which provides precision, recall, and F1-score for each category, as well as a confusion matrix for visualizing the model's performance.

**5. Hyperparameter Tuning:** We performed hyperparameter tuning on the RandomForest and XGBClassifier models using RandomizedSearchCV, which finds the optimal model parameters leading to the most accurate predictions.

**6. Stacking Classifier:** Finally, we build a stacking classifier that leverages the strengths of individual models to improve the overall prediction accuracy. The stacked model includes RandomForest, Bernoulli Naive Bayes, and XGBClassifier.

# Visualization and Evaluation:

The project includes a detailed visualization of the results, showcasing the confusion matrix for each of the models, which gives a clear indication of each model's performance. The evaluation metrics used in this project are precision, recall, and F1-score, which are output as a classification report for each model.

# Future Directions:

While the current results are promising, further work can be done to improve the accuracy of the classification. This includes experimenting with other machine learning models, further tuning the hyperparameters, and testing with larger and more varied datasets.

By harnessing the power of machine learning, this project demonstrates a promising approach to automating consumer complaint classification, leading to more efficient and effective customer service processes.