1. Problem Formulation:

The goal of this project is to use feature selection and classification to determine the main attributes relevant to the predominant (mainhue) color of a country's flag. This can be useful for understanding color trends in flag design and cultural symbolism.

Proposed methods:

Try different sets of attributes and train a classification model based on them. The selection of attributes can be made manually – based on the logical correlation of attributes or programmatically – e.g. iteratively choose every set of 5 attributes or use feature selection methods such as L1 regularization.

2. Dataset:

Source: The dataset used for this project is sourced from http://archive.ics.uci.edu/dataset/40/flags.

Attributes: The dataset includes various attributes related to country flags, such as landmass, zone, area, population, language, religion, and specific color-related features like the presence of red, green, blue, gold, etc.

Size: The dataset comprises 194 instances, providing an entry for every country in the world (in 1990).

The size of the dataset is questionable small, however the goal isn't to train a perfect generalized machine learning model but to use machine learning technics to take a deeper look at the color trends of the word flags.