

Flags Classification

The purpose of this investigation is to see how well a model can predict which landmass a flag belongs to, based on its properties.

Dataset

The dataset the models use is titled: "Flag database" collected by Collins Publishers. Found <https://archive.ics.uci.edu/ml/datasets/Flags>

Examination

The dataset has some values that aren't necessary such as the name, zone, area, population, language and religion as we want to predict how well the properties of the flag and nothing else can be used to predict the landmass. Also, some columns need to be changed to integers values from text. Such as topleft, botright and mainhue. Most of the columns are categorical so a matrix needs to be made for SVM models.

Testing and training split

The database has only 194 entries, so an 80% split should leave the model trained enough for testing.

Results

Model	Training Accuracy (%)	Testing Accuracy (%)
Tree	72.26	43.59
KNN	43.87	53.85
Discriminant Analysis	63.22	30.77
SVM	37.42	10.26
Linear	58.71	48.72

Analysis

Overall models can be used to predict the landmass. The KNN and linear models can predict from a 1 in 6 chance correctly around 50% of the time. SVM performed poorly due to only having 194 entries. I feel like the database doesn't have that much entries and machine learning will therefore not perform that great.

Conclusion

A model can be trained and can be used to predict the landmass, but the limited database ensures that the model doesn't do it that well.