



Original dataset: SharedResponses.csv \*  
Cultures dataset: cultures.csv \*

\* Retrieved at: <https://goo.gl/JXRrBP>

Merge (on='country', how='left')

### Extract and Prepare Sample

Sub-sample (1 per continent)

Data cleaning

Feature engineering

Sub-sample (stratify based on culture)

- Delete unnecessary columns
- Deal with missing values
- Set efficient dtypes
- Delete incomplete scenarios

- A function collects desired features for each scenario and creates a new dataframe

- Reduce size of sample by ensuring that all cultures remain present with similar proportions

x6 (6 continents = 6 sub-samples)

Merge final sub-sample

Visualize distributions

Validation set

Train set

Test set

Prediction with validation set

Train models: Bayesian classifier, Random Forests, SVM, KNN

Hyperparameter tuning

Report results on test set

- Accuracy
- Precision
- Recall
- F1-score