World Happiness Report

Final Project

Mareike Van Wie
Erik Guerra
Marianne Valdespino

Machine Learning

✓ Description of preliminary data preprocessing

✓ Description of preliminary feature engineering and preliminary feature selection, including their decision-making process

✓ Description of how data was split into training and testing sets

✓ Explanation of model choice, including limitations and benefits

Preprocessing and cleaning data

- Dropped columns
- Changed the index to name of countries
- Returned matrix and series of the data using `X = df2.iloc[:,1:]`

  `y = df2.iloc[:, 0]`

- Scaled the data using `scaler = Normalizer().fit(X_train)`

Splitting the data

- The data was split using a test size of point three and randomness of 101
- We then scaled the data and input it into our first model, a Linear Regression model

Models

- We decided to compare three models to see which one is best with our data
- We found the linear regression model was best with a .993 accuracy, decision tree with .992 and random forest with .997
- We chose these three because a. Linear is a good place to start, its advantage is estimation procedure simple and easy to understand 2. Decision trees allow all aspects to be challenged however we also understood it could lead to overfitting of the data and 3. Random forest is quick, allows for high dimensionality and has a low bias

Methods for Visualizations

- Tableau
  - Maps
  - Heat maps
  - Density plots
  - Pie charts

Interactive features we wish to highlight

- Map with clickable urls to reports on countries' happiness levels
- Filter pie charts to see dynamic of each variable (overall)
- Input different variables to see impact on overall happiness measures
  - Have density graphs with ladder + any of six connected variables
- Interactive heatmap that shows impact of variable with hoverable pop ups showing connected country/variable
-

Storyboard

Top Five countries pie chart

Density Plot to show correlation of all variables

Heat Map Displaying all countries (Hovering over each country will show leading variables)

Map with URLs for each countries leading variable that affects happiness