Solving for Happiness

Predicting the Happiness Index of Countries Across Time

Isaac Wecht, Brown University, 12/06/2022
https://github.com/iwecht1/1030project

Recap: Are We Improving?

Population

GDP-per-capita

Life Expectancy

GDP

CO2 Emissions

Recap: Are We Improving?

Population

GDP-per-capita

Happiness Index
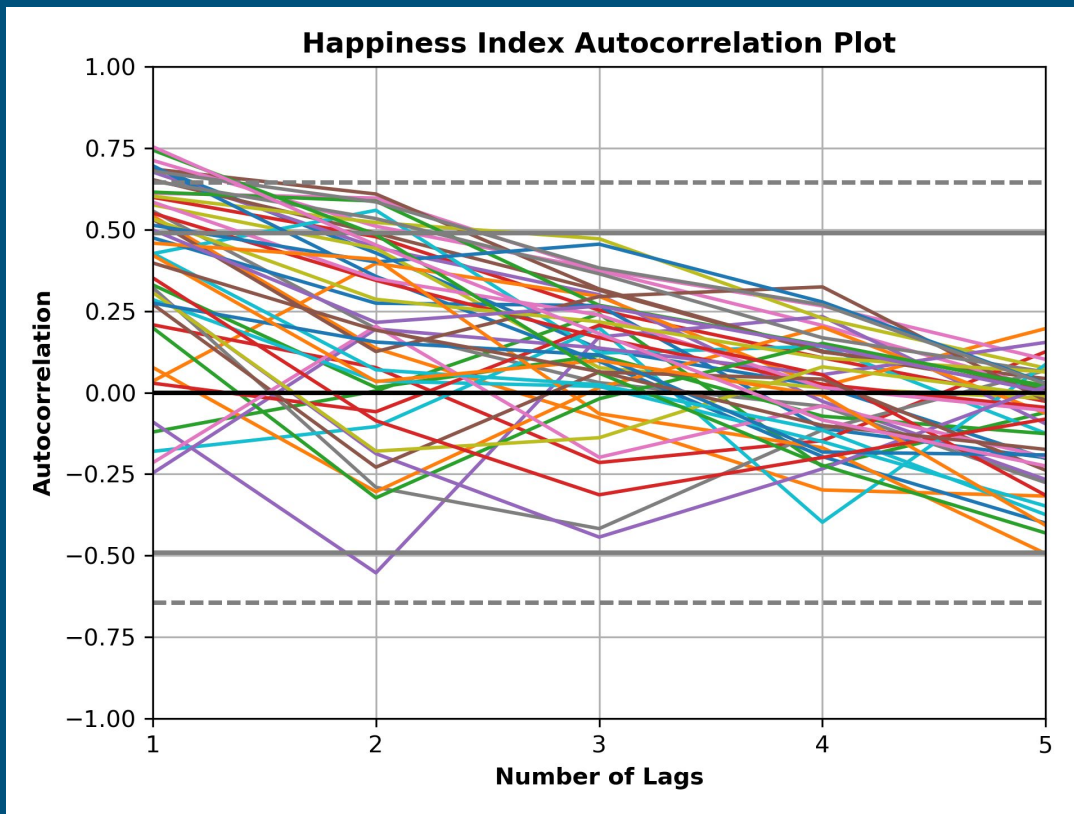
Life Expectancy

GDP

CO2 Emissions

Preprocessing

- OneHotEncoder
  - Feature: Country
  - Result: 47 unique columns for country indicators
- StandardScaler
  - Features: Population, Life Expectancy, GDP, GDP-Per-Capita, CO2 Emissions, Happiness Index

Happiness Index Autocorrelation

Splitting Time-Series Data

- Deterministic Splitting Strategy
  - PredefinedSplit()
  - Time oriented datasets
- Vary Lag Intervals
  - Uncertainty in time lags
  - 5 separate datasets

Country	Target Year	Target Score	Population lag 5 years	Population lag 4 years	Population lag 3 years	Population lag 2 years	Population lag 1 year	Population Current
Argentina	2005	6.073158	NaN	NaN	NaN	NaN	NaN	38892924.0
Argentina	2006	5.961034	NaN	NaN	NaN	NaN	38892924.0	39289876.0
Argentina	2007	6.424133	NaN	NaN	NaN	38892924.0	39289876.0	39684303.0
Argentina	2008	6.441067	NaN	NaN	38892924.0	39289876.0	39684303.0	40080159.0
Argentina	2009	6.775805	NaN	38892924.0	39289876.0	39684303.0	40080159.0	40482786.0

Lagging Trade-offs
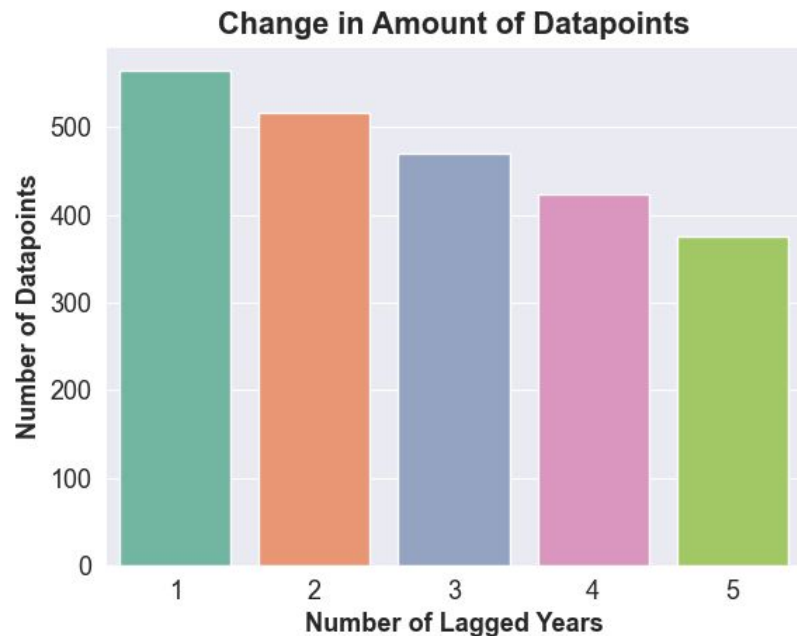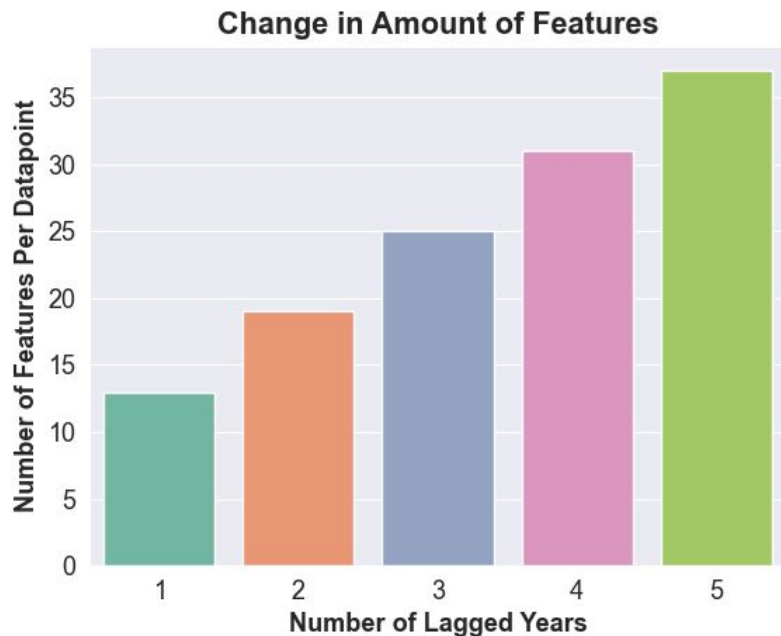
ML Tuning

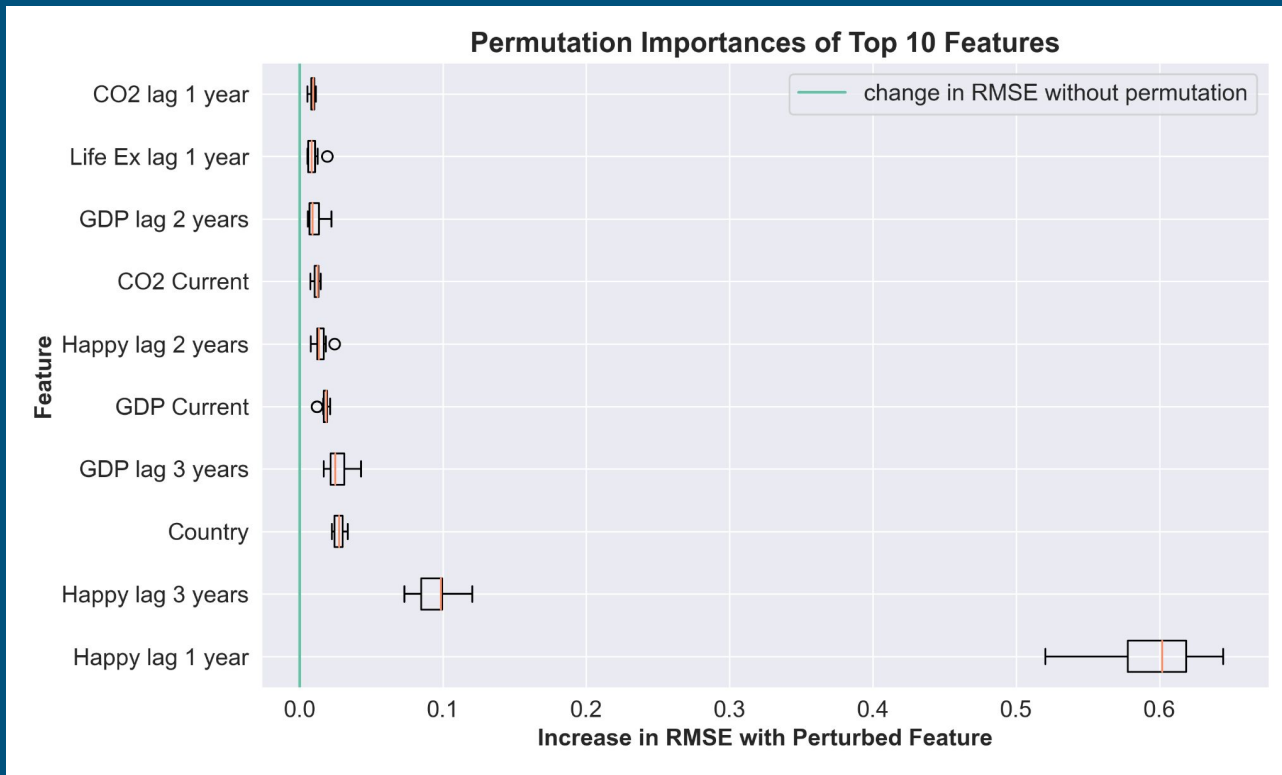|	Model	Hyperparameter(s)	Tested Values
1	Lasso Regression	Alpha	np.logspace(-30,10,21)
2	Ridge Regression	Alpha	np.logspace(-7,7,21)
3	Random Forest	Max Depth Max Features	[1, 3, 5, 7, 10,15,20] [0.15, 0.25, 0.5,0.75,1.0]
4	Support Vector	Gamma C	np.logspace(-8,0,15) np.logspace(-1,3,15)
5	K Nearest Neighbors	Neighbors Weights	np.linspace(1,10,10) ['uniform','distance']

Model Comparison - 0.988 Baseline RMSE

Comparison of Model Test Scores

Ridge Performance

True vs. Predicted Happiness Index
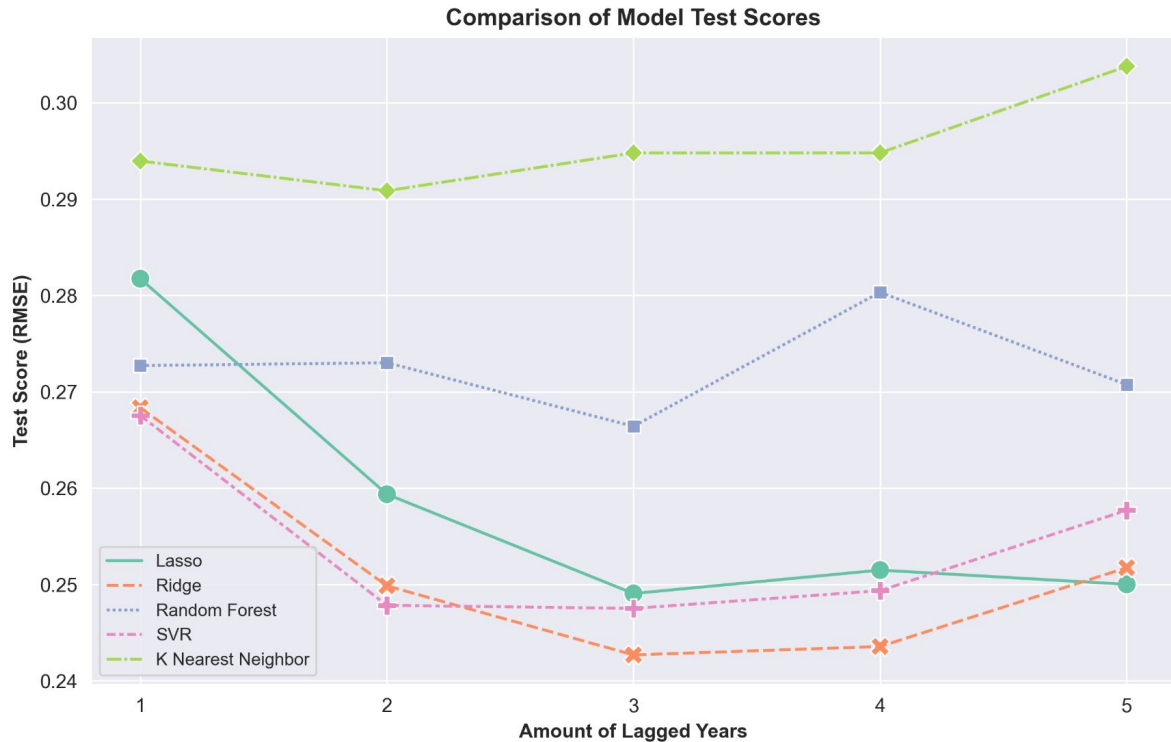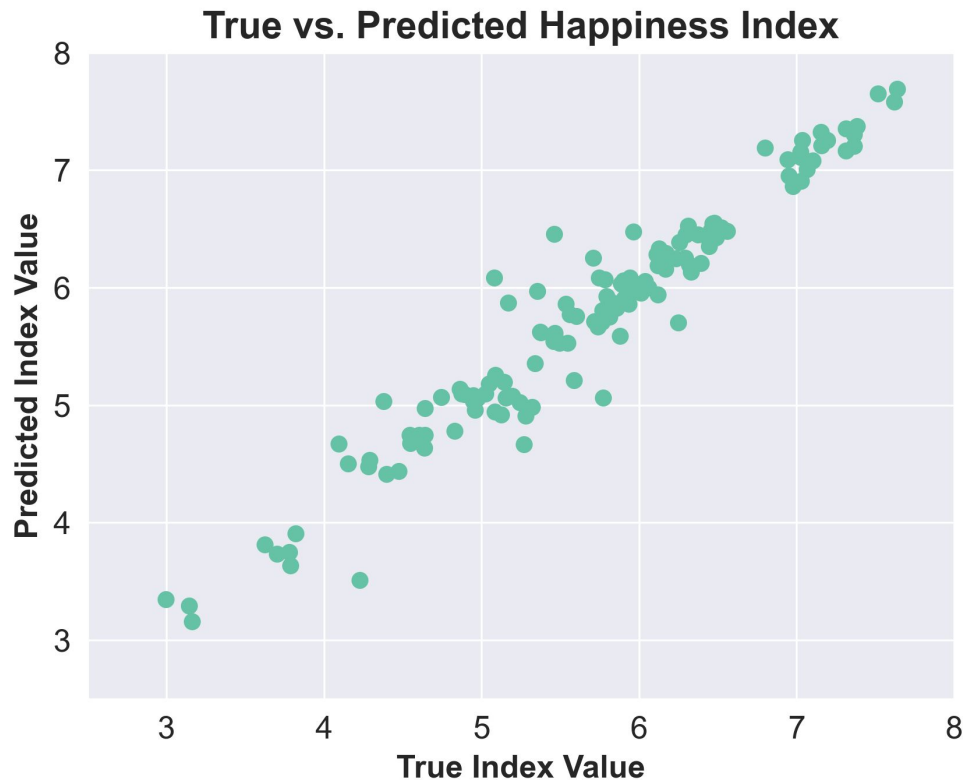
	Model	RMSE	Std
1	Ridge Regression	0.2512	0.0092
2	Support Vector	0.2539	0.0077
3	Lasso Regression	0.2583	0.0122
4	Random Forest	0.2735	0.0149
5	K Nearest Neighbors	0.2956	0.0043

Happiness is a Markov Chain*

*Markov Chains not guaranteed to bring you happiness

Local Feature Importance

higher ⇄ lower

f(x) **5.20**   base value 5.401

4.601   4.801   5.001   5.601   5.801   6.001   6.201

Country_Tanzania | num__Life Ex lag 2 years | num__Happy lag 1 year   cat__Country_Philippines   num__GDP_cap Current | num__Life Ex lag 3 years

Country	Philippines
Target Year	2018

Outlook

- The model can predict the future up until the past
  - Changes in environmental and political realities may weaken model's predictive power
- Narrow scope of happiness
  - Consider bolstering dataset with wider array of environmental, social, and political data

Thank you