

HANNAH LOPEZ  
MAJOR: Computer Science  
Concentration: Cybersecurity

# Happiness Ranking by Country

I thought it would be interesting to see what factors influenced the happiness of a country and people.

## CONTENT ANALYSIS

A brief description of the contents your data file, including interesting descriptive statistics, and any anomalies you had to deal with when loading the file



### Columns:

- GDP per capita
- Generosity
- Life Expectancy
- Freedom to make life decisions
- Social Support

### Added Columns:

- Regional Indicator : One-Hot Encoded (VERY INFLUENTIAL)
- Target (would've been a regression model but I categorized the rankings)

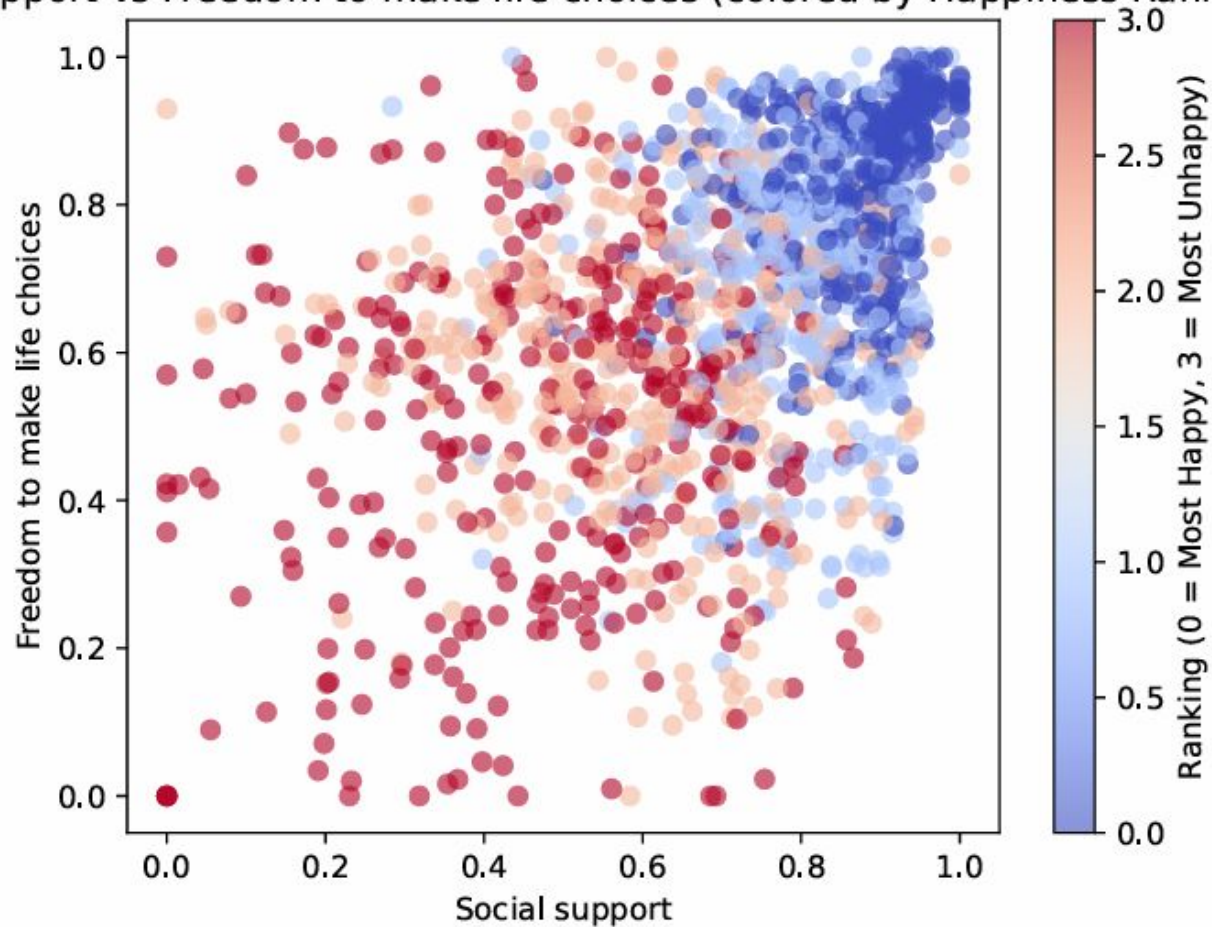
### Descriptive Statistics

- Generosity only increased accuracy by ~2%
- The 2015 dataset had more countries included than any of the other datasets (went up to 2024)
- Most of the happy countries in Western Europe

### Anomalies

- Other dataset was weird
- Perception of Corruption column (replaced with 2015)
- Categorized bins were changed to increase accuracy
- Decimals used commas

Support vs Freedom to make life choices (colored by Happiness Ranking)



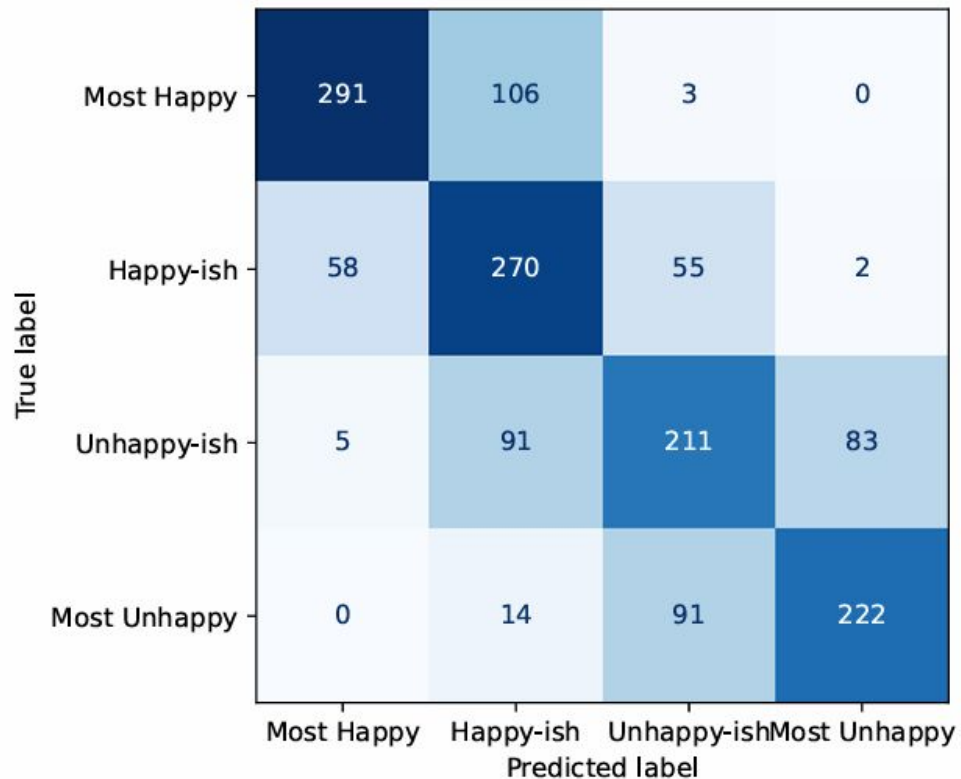
A table listing the raw and balanced accuracy/error for each fold, along with the overall raw and balanced accuracy/error.

Used: Stratified KFold with 10% increments  
And 'rbf' kernel

FOLD	RAW ACCURACY	BALANCED ACCURACY
1	62.3%	62.6%
2	66.9%	66.3%
3	67.3%	67.4%
4	68.7%	69%
5	65.3%	64.6%
6	62%	61.9%
7	67.3%	67.5%
8	70%	70.6%
9	68%	68.2%
10	64%	64%
OVERALL:	66.2%	66.2%

# The Confusion Matrix

W1	W2	W3	W4



## CONCLUSION

### PREVIOUSLY...

My data honestly just made me really mad and frustrated. I had so much but almost all of it was evenly split and led to accuracy levels hovering around the 50% mark. In the end, I would probably look for a different dataset OR if I had time, I would try a few more combinations and see if any of it worked.

### NOW...

I was starting to get a little frustrated with this data too, but then I realized my accuracy was struggling because of the amount of categorical bins I had created. I had to accept that this is as far as I was going to get now, today. If I had more time, I would probably look deeper into what factors correlated better with the target. It overall was really fun!