

Machine Learning Analysis of Food Scarcity and Surplus Through Key Country Statistics



Team 3:

Michael Dasaro, Christopher Morales, Hamza Mustafa,
Abdul Naeem Shaik

Motivations

- ❑ More than 830 million people - roughly 10% of the world's population - goes to sleep hungry at night
- ❑ More than 1.3 billion tons of food are wasted yearly
- ❑ World hunger is not a supply problem!



Problem Formulation

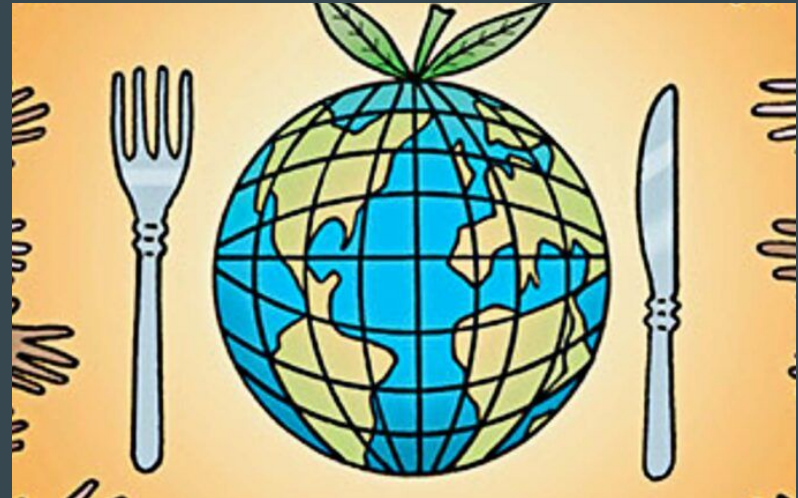
- ❑ Searching for correlations between real world data and hunger
- ❑ Interesting to apply machine learning to such a broad topic
- ❑ Creative data processing
- ❑ Use of multiple Machine Learning models



MARYAM SIKANDER · UPDATED 3 MONTHS AGO

SDG-2: Zero Hunger

SDG-2: Zero Hunger is to create a world free of hunger by 2030, according to UN



Project Setup

- ❑ GitHub: github.com/Michael73MGD/AppliedMachineLearning
- ❑ Jupyter Notebooks for processing and ML
- ❑ Easy division of labor and code merging
- ❑ Easy to write the paper following comments

Michael73MGD rename 52c0335 · yesterday 18 Commits		
Documents	Docs	last week
processed_data	Cleaned up finaltree	2 days ago
raw_data	Data Processing	last week
Machine_Learning_Analysis.ipynb	rename	yesterday
README.md	Data Processing	last week
data_preprocessing.ipynb	Final Dataset with GPS coordinates	5 days ago
requirements.txt	Cleaned up finaltree	2 days ago

Let's load the food scarcity database

```
In [3]: # Loading food scarcity database: Maryam Sikander_Zero_Hunger
# https://www.kaggle.com/datasets/maryamsikander/sdg-2-zero-hunger

# Load the dataframe
scarcity_raw_df = pd.read_excel('raw_data/Maryam Sikander_Zero_Hunger.xlsx', sheet_name="Prevalence-of-food-insecurity")

# Data is provided from years 2015 - 2020. Since our other dataset is recent, we will just take the data from 2020
scarcity_2020_df = scarcity_raw_df[scarcity_raw_df["Year"] == 2020]

# Since year is now always 2020 we can drop the column
scarcity_2020_df = scarcity_2020_df.drop(columns=["Year"])
scarcity_2020_df.head()
```

Raw Data

- ❑ We combined three datasets from Kaggle
- ❑ 226 Countries
- ❑ Food waste (tons/year)
- ❑ Percent population & population → tons of food scarcity

SDG-2: Zero Hunger

SDG-2: Zero Hunger is to create a world free of hunger by 2030, according to UN

Food Waste

Retail, Food Services, and Household Food Waste Estimate of all Countries

World Population Dataset

Global Headcount: World Population Dataset by Country/Territory

Country	Tons of food scarcity	Tons of surplus food
Afghanistan	18414379.0	1852067.5
Albania	597953.0	141775.0
Algeria	5572676.0	2295944.5
Andorra	0.0	3742.5
Angola	17532405.0	1833639.0
Antigua and Barbuda	20641.0	4211.0
Argentina	11247749.0	1971971.5
Armenia	202635.0	160727.0
Aruba	0.0	4610.0
Australia	2061947.0	1400679.0
Austria	198421.0	213269.0
Azerbaijan	659522.0	546010.0

Extended Dataset with Analysis Parameters

❑ Data from The World Bank

- ❑ GDP per Capita
- ❑ Infant Mortality Rate
- ❑ Life Expectancy
- ❑ Population

```
clean_df['normalized_sub'] = clean_df['sub'].apply(lambda x: ((x - min_positive) / (max_positive - min_positive)) * (b - a) + a if x >= 0 else (x - min_negative) / (max_negative - min_negative) * (d - c) + c)
```

❑ Data parsing and normalization

Country	Tons of food scarcity	Tons of surplus food	Coordinates	GDP per Capita	Population	Infant Mortality Rate	Life Expectancy
Afghanistan	18414379	1852067.5	33.7680065, 66.2385139	355.8	41128771	43.4	61.982
Albania	597953	141775	1.2315264, -75.89204331320687	6810.114041	2777689	8.4	76.463
Algeria	5572676	2295944.5	28.0000272, 2.9999825	4342.637968	44903225	19.2	76.377
Andorra	0	3742.5	42.5407167, 1.5732033	41992.77278	79824	2.6	
Angola	17532405	1833639	-11.8775768, 17.5691241	3000.444231	35588987	47.2	61.643
Antigua and Barbuda	20641	4211	17.2234721, -61.9554608	19919.72669	93763	5.2	78.497
Argentina	11247749	1971971.5	-34.9964963, -64.9672817	13650.60463	46234830	6.1	75.39
Armenia	202635	160727	4.536307, -75.6723751	7018.051504	2780469	9.5	72.043
Aruba	0	4610	12.5013629, -69.9618475	33300.83882	106445		74.626
Australia	2061947	1400679	-24.7761086, 134.755	65099.84591	26005540	3.2	83.3

Machine Learning Implementation

- ❑ Checking the fit of each parameter using machine learning methods
- ❑ Decision tree and random forest models were fit
- ❑ A grid search was performed to determine the optimal value of *max leaf nodes* and *n_estimators* for the random forest

```
# Create Tree Classifier
class_tree = tree.DecisionTreeClassifier()
class_tree = class_tree.fit(x_train, y_train)

# Create Forest Classifier
class_forest = rfc(n_estimators=50, max_leaf_nodes=15)
class_forest = class_forest.fit(x_train, y_train)
```

```
from sklearn.model_selection import GridSearchCV as CV
class_GS = CV(estimator=class_forest, param_grid={'n_estimators': n_estimators})
class_GS.fit(x_train, y_train)

print('Best n_estimators: ', class_GS.best_estimator_.get_params()['n_estimators'])

class_scores = class_GS.cv_results_['mean_test_score']
```


Results

ML Model	GDP Per Capita	Infant Mortality Rate	Life Expectancy
Decision Tree	55.81%	58.14%	46.51%
Random Forest	69.77%	58.14%	53.49%

- ❑ The GDP, per capita has a connection with the availability of food as shown by the data. The Random Forest model was most accurate at 69.77% indicating that countries with GDP per capita tend to handle their food resources efficiently.
- ❑ Looking at the Infant Mortality Rate also provided insights. Both the Decision Tree and Random Forest models showed an accuracy of 58.14% hinting at a connection between increased food availability and decreased infant mortality rates.
- ❑ In terms of Life Expectancy it had the accuracy (46.51% for Decision Tree and 53.49%, for Random Forest) suggesting it may not be as directly tied to food scarcity or surplus compared to other factors.

Conclusions

- ❑ We see a strong relationship between GDP per capita and infant mortality rate with food scarcity
- ❑ Life expectancy is not as strongly related – indicates that other factors are more important in determining life expectancy
- ❑ We are glad that we were able to use Machine Learning on global dataset and draw conclusions

References

Links to Data Used:

- ❑ GDP per Capita: <https://data.worldbank.org/indicator/NY.GDP.PCAP.CD>
- ❑ Infant Mortality Rate: <https://data.worldbank.org/indicator/SP.DYN.IMRT.IN>
- ❑ Life Expectancy: <https://data.worldbank.org/indicator/SP.DYN.LE00.IN>
- ❑ Population: <https://data.worldbank.org/indicator/SP.POPTOTL>

Additional Resources:

- ❑ <https://www.actionagainsthunger.org/the-hunger-crisis/world-hunger-facts/>
- ❑ <https://www.thenation.com/article/world/world-hunger/>
- ❑ <https://reason.com/2022/04/23/californias-new-food-waste-law-is-backfiring/>
- ❑ <https://www.weforum.org/agenda/2018/01/internet-things-iot-world-hunger-supply-chain/>