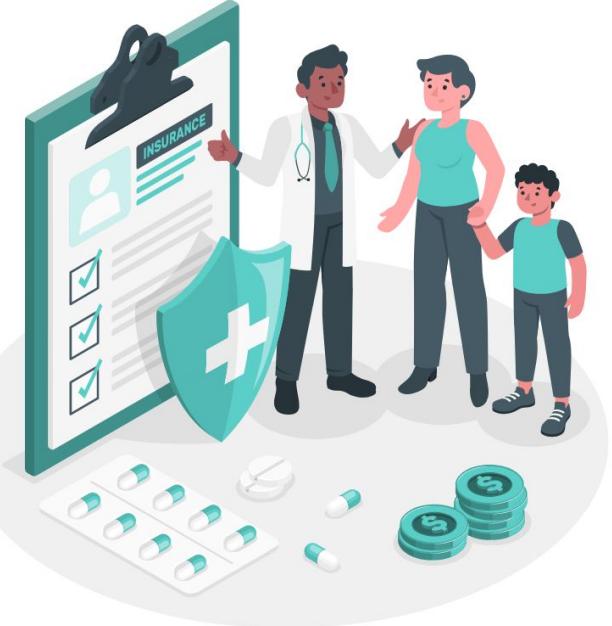


# UN Sustainable Development Goal Text Classification

By. Suchanya (Mild) Trakarnsakdikul



# TABLE OF CONTENTS



**01**

## United Nation

What are the Sustainable Development Goal



**02**

## The Problem

The problem and goal for this project



**03**

## The Data

Getting to know and seeing the data



**04**

## Modeling

Explore the different model used in this project



**05**

## Recommendation

What was the best model and how it helps the problem



**06**

## The Future

What can be done more to take the solution futher



# What are the Sustainable Goals?

The United Nation created 17 measurable indication for development refer to as Sustainable Development Goals or refer to as SDG that they hope to achieve by 2030.



# Our Target



## Goal 3: Good Health and Well Being

- 27 Target or indicators that provide quantitative measurement
- The indicators will be used to judge whether the SDG has been achieved in 2030 or not

# The Problem

Since there are many indicators, the UN, donors and developers struggle to identify which tenders, programs or news articles are related to which 27 indicators.

**Our Goal: Develop a text classifier to identify which tenders are most relevant to which indicators**

# The Data



## The Train Data

Includes **2995** web-scraped text with:

- **Unique ID:** IDs of text to be classified
- **Type:** Type or source of text
  - Grants      ○ Program
  - Tenders     ○ Organization
  - Contract    ○ Open opportunity
  - News        ○ Funding information
- **Text:** text to be classified
- **Labels:** Indicators for each text (ie 3.1.1)

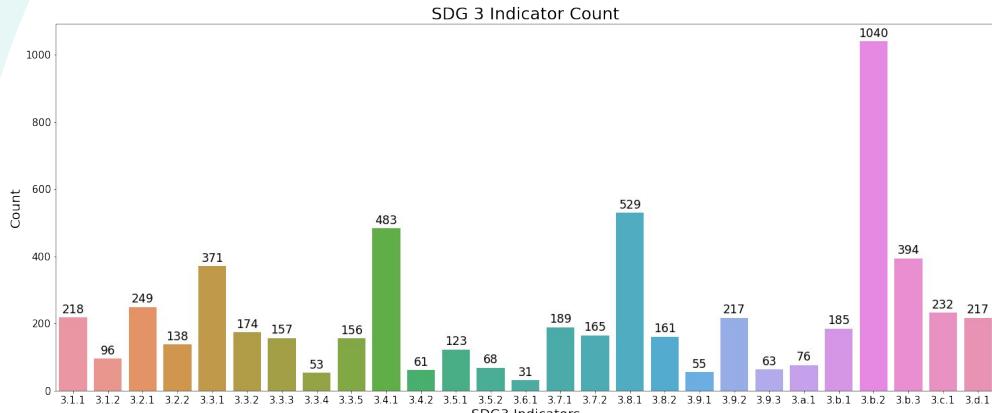


## The Test Data

Includes **998** web-scraped text with:

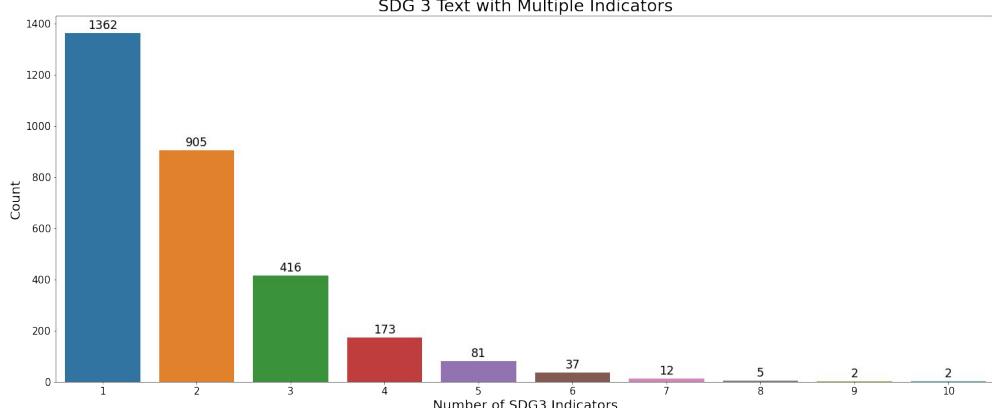
- **Unique ID:** IDs of text to be classified
- **Type:** Type or source of text
  - Grants      ○ Program
  - Tenders     ○ Organization
  - Contract    ○ Open opportunity
  - News        ○ Funding information
- **Text:** text to be classified

# Understanding the Indicators



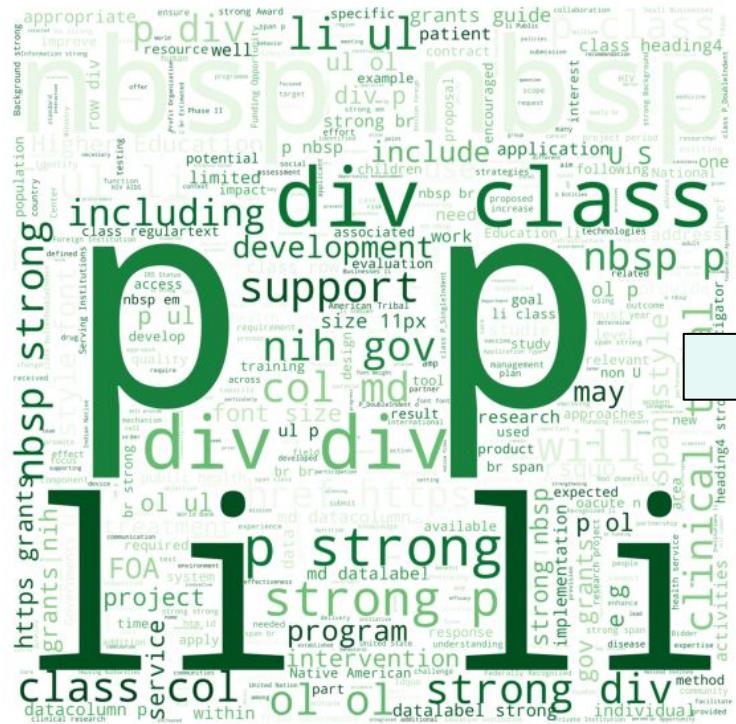
27 Indicators:

- Most labeled indicator:
  - 3.b.2 - International Health Regulation
  - 3.8.1 - Coverage of essential health services
- Most text have 1 to 2 indicators

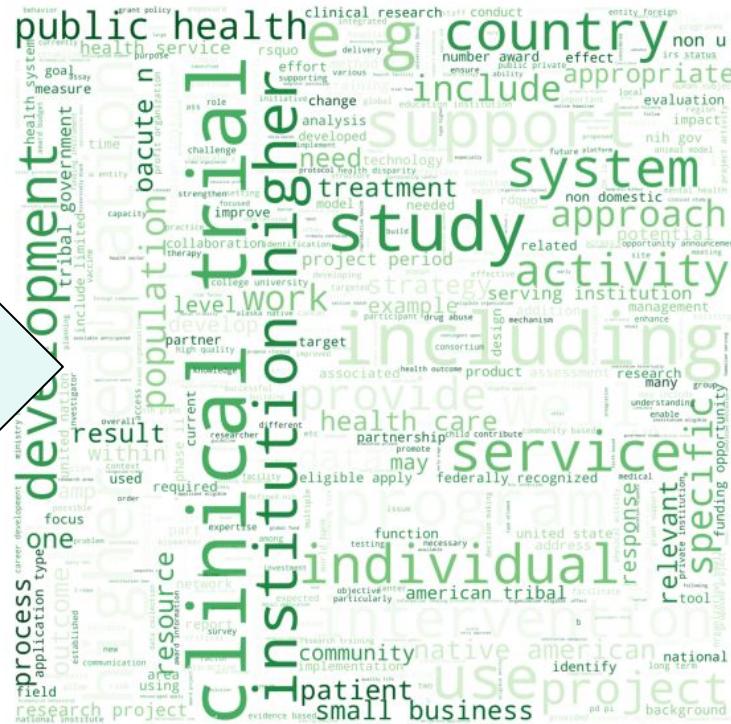


## Data at a Glance

# Original Data



## Cleaned Data

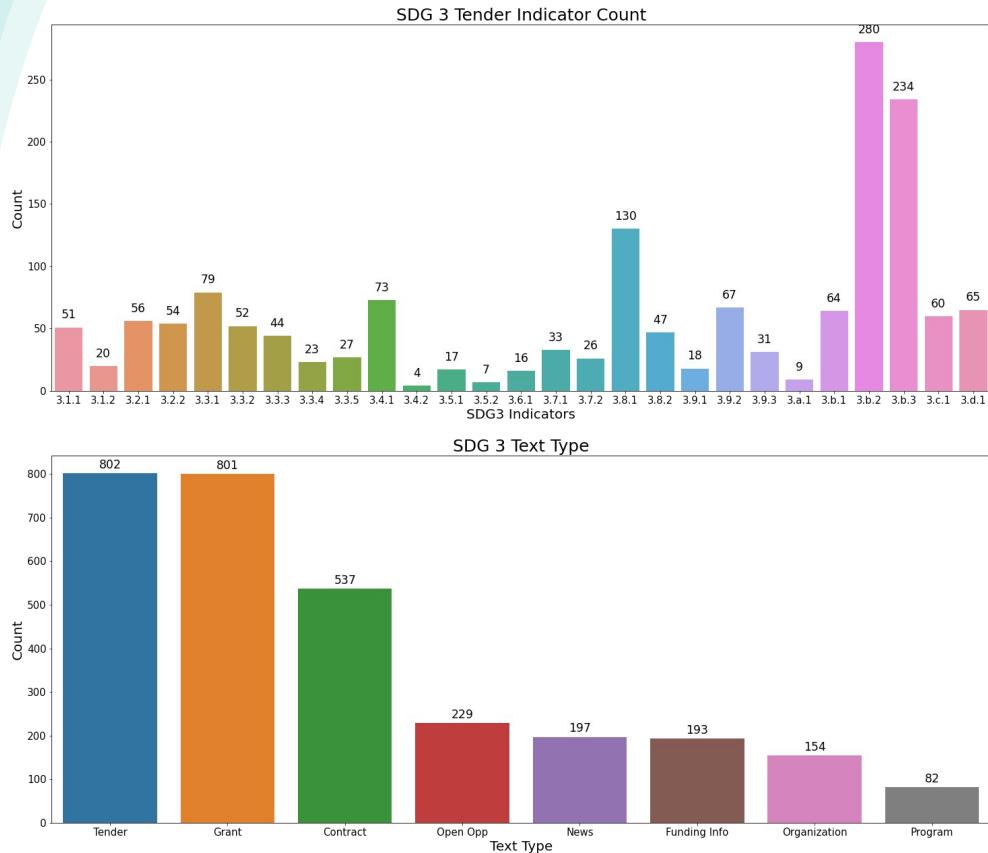


# Indicator Examples

## Indicator 3.2.1: Under-5 mortality rate

## Indicator 3.3.2: Tuberculosis incidence

# Understanding the Text Type



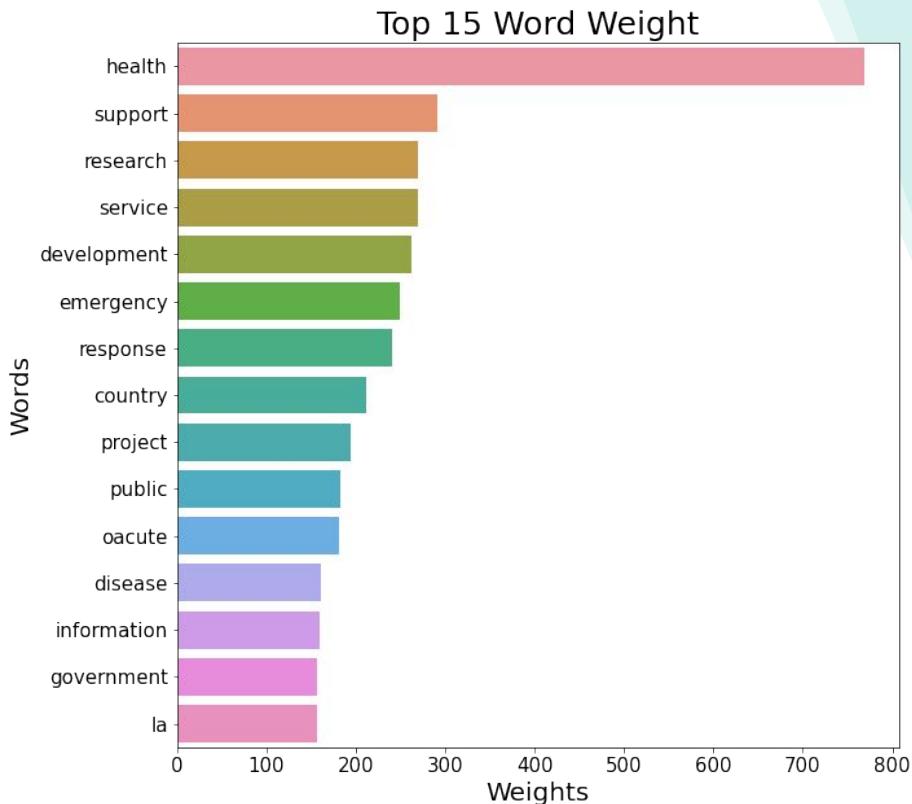
## 8 Text Type:

- By Description
  - Grant
  - Organization
  - Tender
  - Funding Information
  - Open Opportunity
  - Program
- By Title or Headline
  - Contract
  - News

# Modeling and Prediction

## KNN, MLP and Ridge Classifier

- TF-IDF Vectorizer to determined the value and weight of each words
- Fitted the best parameters for the estimators and vectorizers
- From the fitted parameter, an optimized model is created with high accuracy





# Best Model and Recommendation

## Ridge Classifier CV

- Use with MultiOutputClassifier function
- Use TF-IDF Vectorizer and ridge classifier default parameters
- 4.47% of the indicators were incorrectly predicted

# Evaluation

8 Text Type:

- **By Description**
  - Grant
  - Organization
  - Tender
  - Funding Information
  - Open Opportunity
  - Program
- **By Title or Headline**
  - Contract
  - News

# Additional Data

8 Text Type:

- **By Description**
  - Grant
  - Organization
  - Tender
  - Funding Information
  - Open Opportunity
  - Program
- **By Title or Headline**
  - Contract
  - News

# Thank you

Appreciate your time



# Appendix

## The Indicators

- The code and what it is

## Data Visualization

- Word cloud for each indicator
- Indicator spread for type of data

# 27 Indicators

There are 27 possible SDG 3 indicators, in the dataset each indicator will be refer using code, e.g.

"3.1.1":

- 3.1.1 - Maternal mortality ratio
- 3.1.2 - Proportion of births attended by skilled health personnel
- 3.2.1 - Under-5 mortality rate
- 3.2.2 - Neonatal mortality rate
- 3.3.1 - Number of new HIV infections per 1 000 uninfected population, by sex, age and key populations
- 3.3.2 - Tuberculosis incidence per 100 000 population
- 3.3.3 - Malaria incidence per 1 000 population
- 3.3.4 - Hepatitis B incidence per 100 000 population
- 3.3.5 - Number of people requiring interventions against neglected tropical diseases

# 27 Indicators

- 3.4.1 - Mortality rate attributed to cardiovascular disease, cancer, diabetes or chronic respiratory disease
- 3.4.2 - Suicide mortality rate
- 3.5.1 - Coverage of treatment interventions (pharmacological, psychosocial and rehabilitation and aftercare services) for substance use disorders
- 3.5.2 - Harmful use of alcohol, defined according to the national context as alcohol per capita consumption (aged 15 years and older) within a calendar year in litres of pure alcohol
- 3.6.1 - Death rate due to road traffic injuries
- 3.7.1 - Proportion of women of reproductive age (aged 15–49 years) who have their need for family planning satisfied with modern methods
- 3.7.2 - Adolescent birth rate (aged 10–14 years; aged 15–19 years) per 1 000 women in that age group

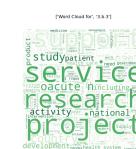
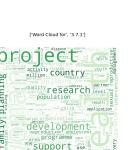
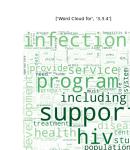
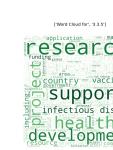
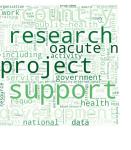
# 27 Indicators

- 3.8.1 - Coverage of essential health services (defined as the average coverage of essential services based on tracer interventions that include reproductive, maternal, newborn and child health, infectious diseases, non-communicable diseases and service capacity and access, among the general and the most disadvantaged population)
- 3.8.2 - Proportion of population with large household expenditures on health as a share of total household expenditure or income
- 3.9.1 - Mortality rate attributed to household and ambient air pollution
- 3.9.2 - Mortality rate attributed to unsafe water, unsafe sanitation and lack of hygiene (exposure to unsafe Water, Sanitation and Hygiene for All (WASH) services)
- 3.9.3 - Mortality rate attributed to unintentional poisoning

# 27 Indicators

- 3.b.2 - Total net official development assistance to medical research and basic health sector
- 3.b.3 - Proportion of health facilities that have a core set of relevant essential medicines available and affordable on a sustainable basis
- 3.c.1 - Health worker density and distribution
- 3.d.1 - International Health Regulations (IHR) capacity and health emergency preparedness
- 3.a.1 - Age-standardized prevalence of current tobacco use among persons aged 15 years and older
- 3.b.1 - Proportion of the target population covered by all vaccines included in their national programme

# Word Cloud



# Indicator Distribution by Text Type

