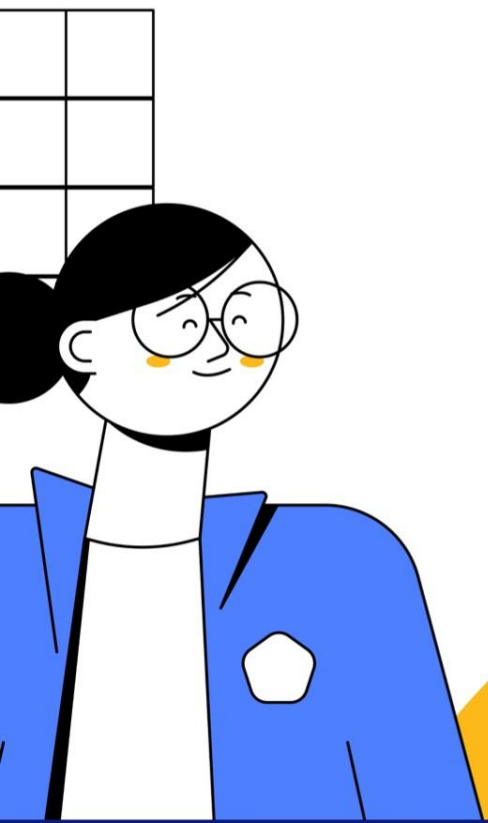
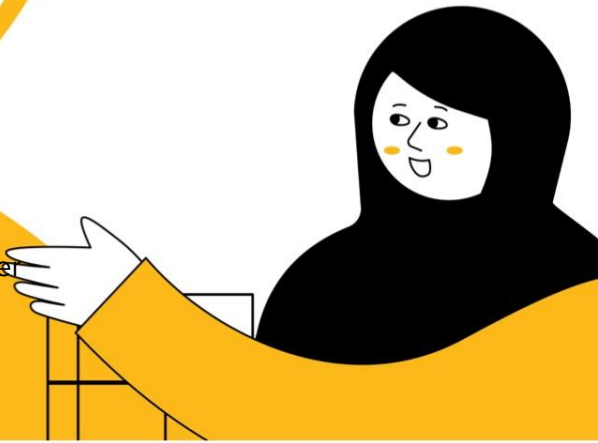
Generasi Gigih

Powered by YAYASAN ANAK BANGSA BISA goto

# Employability


Ichlas Magribi_KM_G2DA5014, Project Manager
Rafi Naufal Al Mochtari Pohan_P_G2DA2062, Technical Research
Syifa Auliyah Hasanah_KM_G2DA9297, Research Design
Chandra Dzaki Kusuma_KM_G2DA2065,
Asiyah_P_G2DA3127,

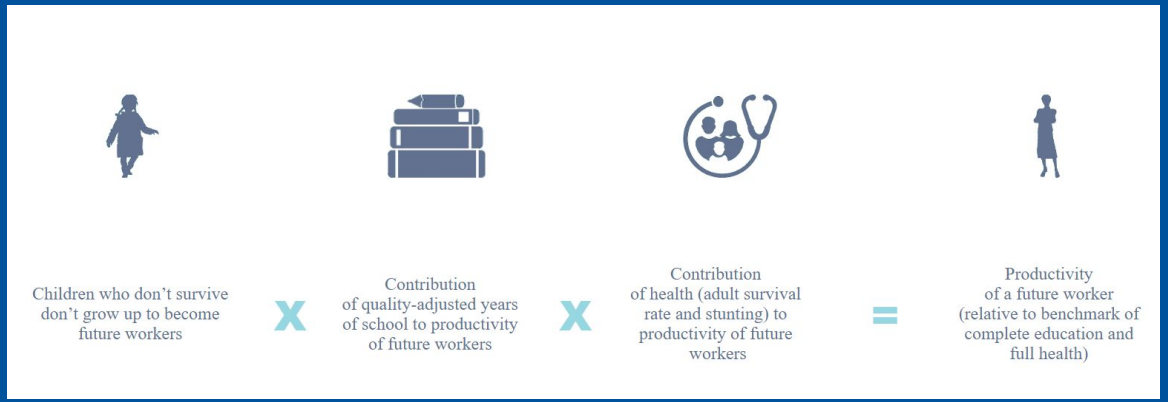


# Table Of Contents



# Human Capital Index

Briefly, **Human Capital Index** is the index shows productivity as a future worker of a child born in that year.




**Let's say country X get 0.5 HCI, what is 0.5 HCI means?**
Every child born in this country has 0.5 chance of growing up provided they complete their education and are in good health.

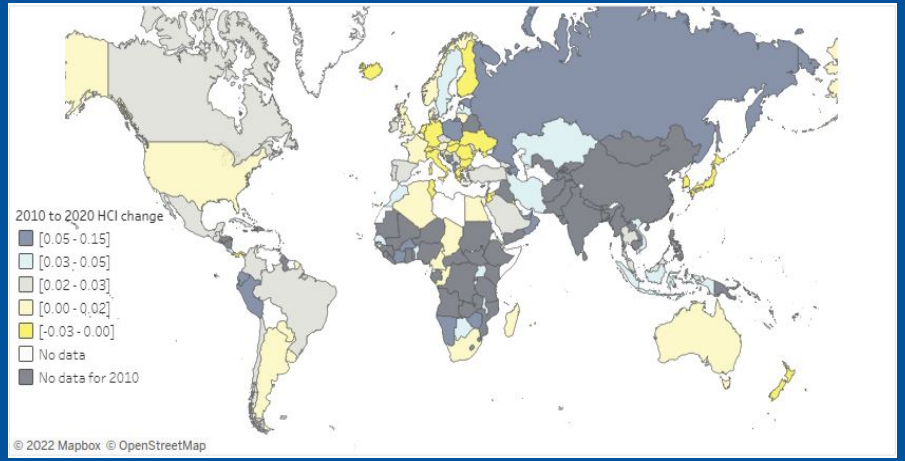
**FYI:** World Bank only released Human Capital Index (HCI) for 4 times (2010, 2017, 2018, 2020), so the data used for the analysis is limited to those 4 years.

# Problem Statement

**How would the Indonesian Government increase its Human Capital Index to 0.64 (high level) to achieve Indonesia Emas by 2045?**

**Human Capital Index 2020**


**HCI 2010 to HCI 2020 Comparison**




- Indonesian HCI is still in the lower middle level
- For 10 years (2010 to 2020), Indonesia HCI only increased 0.04
- Indonesia's HCI score (0.54) is still lower than the average HCI score in ASEAN (0.58)

# ISSUE TREE DIAGRAM

**How to increase HCI?**

- Survival
  - Probability of Survival to Age 5 (0-1)
- Education
  - Expected Years of School (0-14)
  - Harmonized Test Score (300-625)
- Health
  - Fraction of Children Under 5 Not Stunted (0-1)
  - Fraction of 15-Year Olds Who Survive to Age 60 (0-1)

| HCI Calculation Formula | Description |
| --- | --- |
| Probability of Survival to Age 5 (0-1) | Calculated by subtracting the under-5 mortality rate from 1. |
| Expected Years of School (0-14) | Calculated as the sum of age-specific enrollment rates between ages 4 and 17. |
| Harmonized Test Score (300-625) | Based on major international student achievement testing programs measured in TIMMS-equivalent units |
| Fraction of Children Under 5 Not Stunted (0-1) | Calculated by subtracting suntuing rates from 1. |
| Fraction of 15-Year Olds Who Survive to Age 60 (0-1) | Calculated by subtracting thr mortality rate for 15-60 year-olds from 1. |

# ISSUE TREE DIAGRAM

**How to increase HCI?**

- Survival
  - Probability of Survival to Age 5 (0-1)
- Education
  - Expected Years of School (0-14)
  - Harmonized Test Score (300-625)
- Health
  - Fraction of Children Under 5 Not Stunted (0-1)
  - Fraction of 15-Year Olds Who Survive to Age 60 (0-1)

**HCI Calculation Formula**

**But how would government controls them?**

**These variables seems depend on another factors and hard to be controlled**

# ISSUE TREE DIAGRAM

**How to increase HCI?**

- Survival
- Education
- Health

**HCI Calculation Formula**

- Probability of Survival to Age 5 (0-1)
- Expected Years of School (0-14)
- Harmonized Test Score (300-625)
- Fraction of Children Under 5 Not Stunted (0-1)
- Fraction of 15-Year Olds Who Survive to Age 60 (0-1)

**Latent factors based on literature study**

- Pneumonia death <5 years
- Diarrea death <5 years
- External Debt Stock
- Public spending on education
- Income (GDP)
- Parental education
- Maternal health access service
- Family planning
- Sanitation condition
- Number of death by accidents
- Public spending on health

**Available data for predictive model**

All data are collected from World Bank and WHO database

# Raw Data

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | **Country** | **HCI** | **EDU** | **HEALTH** | **SANI** | **DEBT** | **GDP** | **PNEMO** | **DIARE** |
| 2 | Aruba | | 5.924789906 | | | | 24631.1821 | | |
| 3 | Afghanistan | | | 9.81848717 | | 2480214114 | 437.26874 | | |
| 4 | Angola | | | 3.84260869 | | 20277281253 | 3122.7816 | | |
| 5 | Albania | 0.54360294 | | 4.6022253 | 43.8438649 | 4605163642 | 4114.1349 | 94 | 1.356009824 |
| 6 | Andorra | | 3.142920017 | 6.23957682 | 79.2656587 | | 43504.2157 | | |
| 7 | United Arab Emirates | 0.62127405 | | 4.04999352 | 79.6917076 | | 32024.1978 | 12 | 1.045706488 |
| 8 | Argentina | 0.58855706 | 5.531050205 | 9.45598793 | 49.756298 | 133695000000 | 8225.13758 | 757 | 5.173022196 |
| 9 | Armenia | | 3.840790033 | 8.44377995 | 52.4277415 | 4934959250 | 2994.34047 | | |
| 10 | American Samoa | | | | | | 11910.0132 | | |
| 11 | Antigua and Barbuda | | 2.508889914 | 4.19123125 | | | 14160.562 | | |
| 12 | Australia | 0.75484377 | 5.085820198 | 8.56316471 | 63.6234566 | | 42783.3226 | 37 | 0.395517581 |
| 13 | Austria | 0.73778212 | 5.730249882 | 10.2258253 | 99.6489704 | | 48153.324 | 6 | 0.394432369 |
| 14 | Azerbaijan | 0.49745739 | 3.224309921 | 2.6137836 | 24.5081648 | 4548547550 | 4950.29479 | 2449 | 18.71502521 |
| 15 | Burundi | 0.34287551 | 6.171949863 | 10.3502188 | | 607230101.8 | 212.137057 | 3474 | 394.6345388 |
| 16 | Belgium | 0.75272685 | 6.458099842 | 10.3545504 | 79.131369 | | 44760.2912 | 12 | 0.897797304 |
| 17 | Benin | 0.36558479 | 3.089270115 | 2.82528329 | | 1325336572 | 1088.75791 | 6003 | 207.6609991 |
| 18 | Burkina Faso | 0.31972128 | | 5.08159304 | | 1939710439 | 624.175165 | 13557 | 367.1155504 |
| 19 | Bangladesh | | 1.939419985 | 2.56097794 | 27.3595642 | 25377795050 | 702.26442 | | |
| 20 | Bulgaria | 0.63661927 | 4.274290085 | 6.59577608 | 51.4179341 | 55682965449 | 6988.23332 | 167 | 3.95242769 |
| 21 | Bahrain | 0.60491228 | | 4.05684614 | 79.0974583 | | 19355.921 | 5 | 1.834889025 |
| 22 | Bahamas, The | | | 5.15085459 | | | 28552.5172 | | |
| 23 | Bosnia and Herzegovina | | | 9.02212143 | 22.5939087 | 14005774333 | 4714.69375 | | |

# Descriptive Statistics

```
#Descriptive Statistics
print('Descriptive Statistics pada Data :')
print(df.describe())
print('=====================================\n')
```

```
Descriptive Statistics pada Data :
                HCI         EDU      HEALTH        SANI          DEBT
count   601.000000  601.000000  601.000000  601.000000  6.010000e+02
mean      0.569461    4.488071    6.679110   61.626751  4.580971e+10
std       0.145609    1.489106    2.796722   23.737210  1.497183e+11
min       0.286075    1.326320    2.083311    5.866489  8.090757e+07
25%       0.443490    3.729640    4.512254   49.756298  7.890234e+09
50%       0.574536    4.366525    6.470227   64.581756  1.284940e+10
75%       0.690000    4.951570    8.458914   77.152274  1.857658e+10
max       0.887084   12.903480   23.961813  100.000000  2.114160e+12

                GDP         PNEMO        DIARE
count    601.000000    601.000000   601.000000
mean   15154.805095   3728.326955    61.500267
std    20096.511200  15124.142457   131.840417
min      212.137057      0.000000     0.135771
25%     2169.991500     19.000000     1.005310
50%     6119.762340    147.500000     5.173022
75%    19133.757800   2279.000000    55.155620
max   113218.713000  164659.000000  1030.029386
=====================================
```

9 Columns x 869 rows, all null cells imputed with median values of each columns

# Project Goals

Provide insights and prediction model to the government in order to improve human capital

# Success Metrics

Get the best prediction model accuracy value above 80%

Presents a visualization that can help the government to know the condition of HCI Indonesia

1. Provide factual information related to indicators of HCI

2. Assist the government in making new policies that can increase the Human Capital Index (HCI) in Indonesia.

# How the product solve the problem

# Tech stack of the product

## STEP 1: PRE-PROCESSING

Google Sheets

Data collection and analysis

Pre-processing data and descriptive statistics

## STEP 2: EXPLORATORY DATA ANALYSIS

Google Data Studio

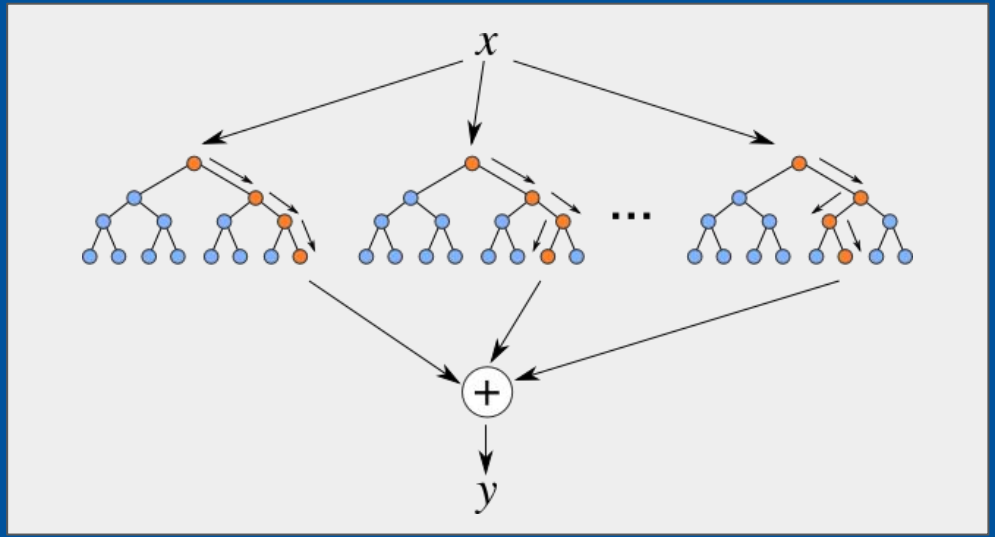
Data visualization

# Tech stack of the product (2)

## STEP 3: MODELLING AND ANALYSIS

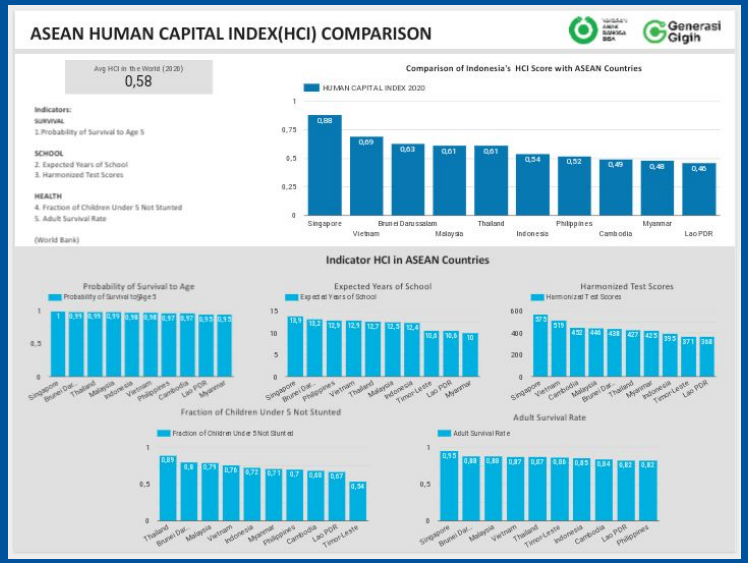
Predictive model visualization (integrated with Python), implement machine learning model- Random Forest Regression Algorithm to predict HCI

## Random Forest Regression Algorithm (Machine Learning Model)

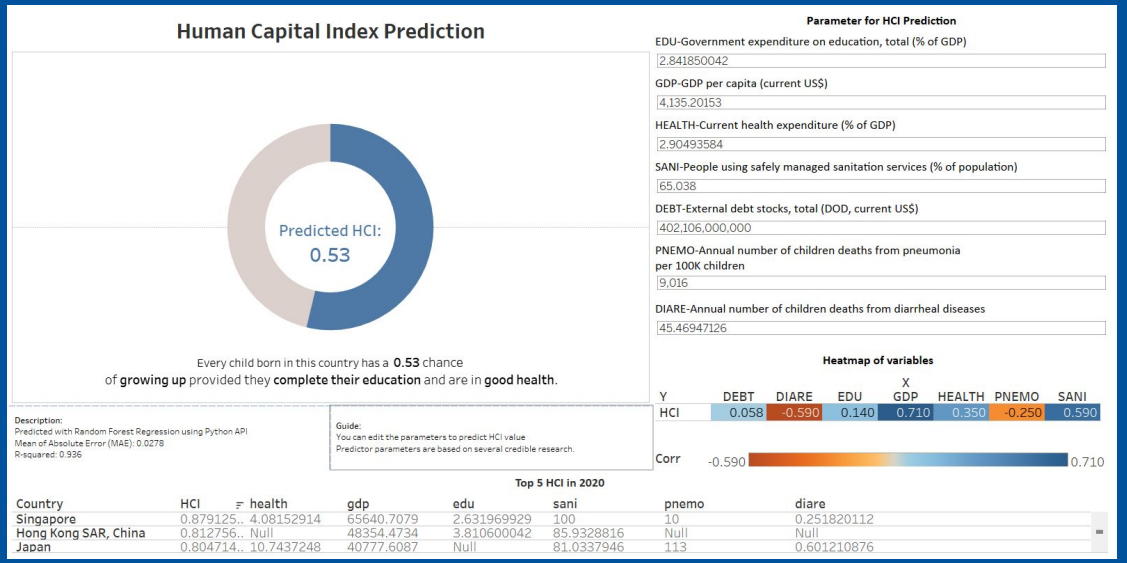


# DEMO TIME!

## Google Data Studio




## Tableau + Python



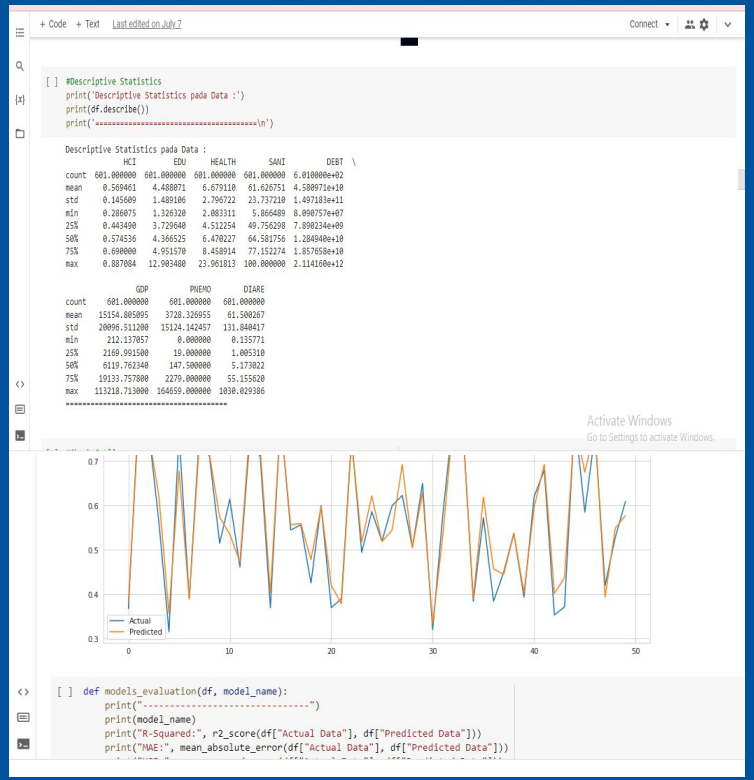

Click here

# Learning Takeaways

- Communication is a key for teamwork!
- Time management is essential for project management!

# Documentation

## Robust Python Model



```
[ ] #Descriptive Statistics
    print('Descriptive Statistics pada Data :')
    print(df.describe())
    print('=====================================\n')

    Descriptive Statistics pada Data :
                  HCI         EDU      HEALTH        SANI          DEBT  \
    count  601.000000  601.000000  601.000000  601.000000  6.010000e+02
    mean     0.569461    4.488071    6.679110   61.626751  4.580971e+10
    std      0.145609    1.489106    2.796722   23.737210  1.497183e+11
    min      0.286075    1.326320    2.083311    5.866489  8.090757e+07
    25%      0.443490    3.729640    4.512254   49.756298  7.890234e+09
    50%      0.574536    4.366525    6.470227   64.581756  1.284940e+10
    75%      0.690000    4.951570    8.458914   77.152274  1.857656e+10
    max      0.887084   12.903480   23.961813  100.000000  2.114160e+12

                    GDP         PNEMO        DIARE
    count    601.000000    601.000000   601.000000
    mean   15154.805095   3728.326955    61.500267
    std    20096.511200  15124.142457   131.840417
    min      212.137057      0.000000     0.135771
    25%     2169.991500     19.000000     1.005310
    50%     6119.762340    147.500000     5.173022
    75%    19133.757800   2279.000000    55.155620
    max   113218.713000  164659.000000  1030.029386
    =====================================
```






```
[ ] def models_evaluation(df, model_name):
        print("------------------------------")
        print(model_name)
        print("R-Squared:", r2_score(df["Actual Data"], df["Predicted Data"]))
        print("MAE:", mean_absolute_error(df["Actual Data"], df["Predicted Data"]))
```

## Solid Team

# Thank you!

# ATTACHMENTS

BLANK INTENTIONALLY

# Analysis and Recommendations

# Human Capital Index

**Analysis :**

- The value of each indicator influences the HCI value (Probability of survival to age 5, expected years of school, harmonized test score, fraction of children under 5 not stunted, adult survival rate). Countries with high scores for each indicator will also produce high HCI scores. For example, Singapore.
- The value of Indonesia's HCI continues to increase from 2010 to 2020, which is also influenced by the increase in the value of each indicator from year to year.

# Probability Of Survival to Age 5

**Analysis :**

- The probability of survival to the age 5 indicator seems to increase from 2010 to 2018. However, in 2020 the value tends to be the same. The value of the indicator rise is impacted by the variable death rate diarrhea and pneumonia. Where the death rate for certain diseases tends to decrease from year to year which boosts the indicator value.

# Education

## Analysis:

- Indonesia's External Debt Stock is the highest among ASEAN countries
- Although Indonesia's External Debt Stock experienced a large increase, it only had a slight increase in the Expected year of school variable in Indonesia
- Harmonized test scores in Indonesia tend to be the same from 2010-2020
- GDP in Indonesia is still lower than the average GDP in ASEAN
- Even though expenditure on education in Indonesia is higher than Singapore, the harmonized test score is still lower than Singapore because Singapore's GDP is higher.

# Fraction Of Children Under 5 Not Stunted

## Analysis :

- Indicators of the number of children under five being stunted are influenced by several variables, including sanitation, parental education, access to maternal health care, and family planning. The value of these indicators tends to improve every year. From 2018 to 2020, there was a fairly high increase. This was attributable to an increase in the quality of home sanitation, enhanced maternal education, and also high service coverage, even if the number of family planning users was falling.

# Fraction Of 15 Year Olds Who Survive To Age 60

**Analysis :**

- The pattern of causes of death in the age range 15–44 years is traffic accidents, whereas for fatalities in the age group 45–54 years are many illnesses. The fraction indicator value tends to remain steady every year due to the resilience of accidents and fatalities each year. In addition, the value of health care does not problem is evident from year to year.

# Prediction Analysis

## Analysis :

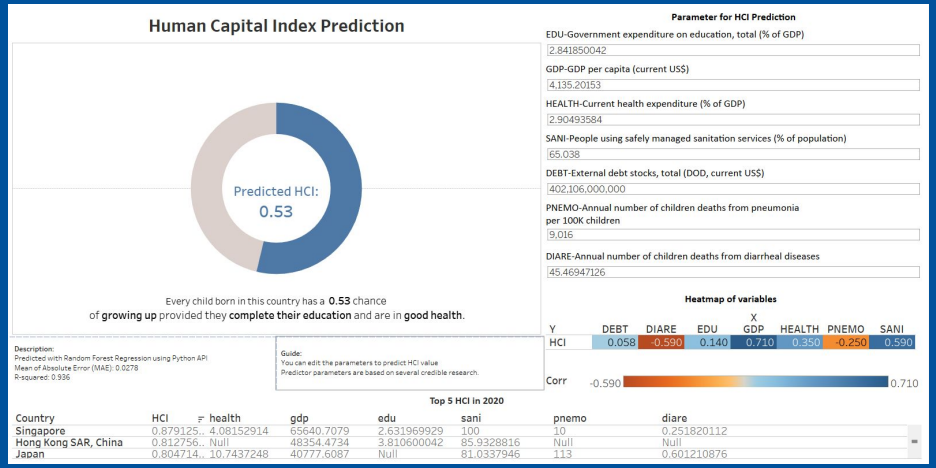
- We use Random Forest Regression to obtain a model that can be used to predict HCI values with several influencing factors. Previously, we analyzed the correlation of each factor to the HCI value and found that the GDP, sanitation, and health spending variables had a higher correlation than the other variables. The model is obtained with an MAE evaluation value of 0.028, which means that the model can predict the HCI value with an error probability of only 2.8%. The model has been very well used.
- An analysis was conducted to determine which variable value predictions should be improved so that HCI Indonesia can achieve the desired HCI target number, which is close to the value of 80%. The result is that the GDP value with the highest correlation will help increase HCI quickly, followed by the value of sanitation, health expenditure, and education expenditure.

# Prediction Analysis

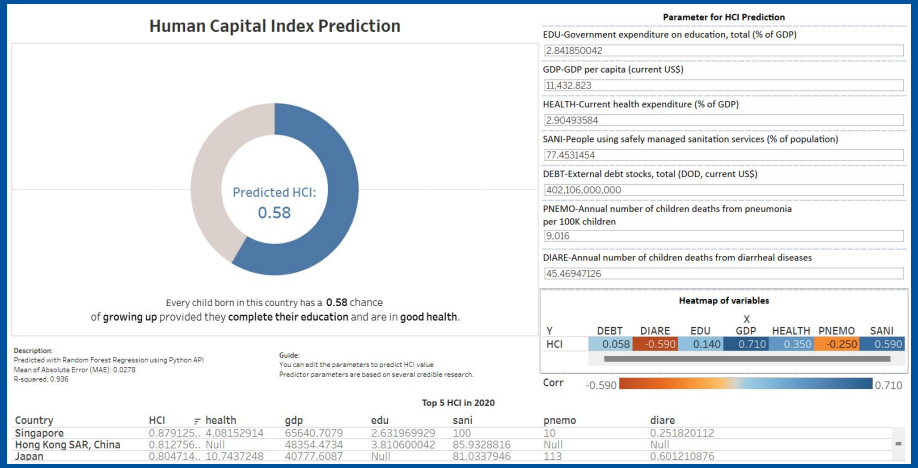
## Analysis :

We tried to make changes to the variables with a high correlation value to the HCI value (GDP and Sanitation), and the results obtained were significantly able to increase the HCI value.

Prediction results for Indonesia's HCI value 2020




Prediction results for Indonesia's HCI value 2020 (with changes)



# Recommendations

- Quality of household condition is the highest correlated controllable variables, then **creating a program to continuously improve the quality of household sanitation**, for example is CLTS program by Yayasan DIan Desa Yogyakarta. CLTS (Community-Led Total Sanitation). CLTS is an approach method to change awareness by initiating or triggering the community's disgust and shame about sanitary conditions where they defecate in the open. So, in the end, they find a joint solution to change conditions.
- **Increase the education budget** so that the quality of education in Indonesia improves. Improve the quality of teachers, education, and school facilities in every region in Indonesia and develop educational curriculum that can compete in the digital era.
- **Increase the spending in the health sector to increase the quality of health services**, reduce stunting in children, provide education, and health insurance for the people of Indonesia.