

EXECUTIVE SUMMARY

U.N SDG text classification task



● Motivation

- Increase the productivity of UN SDG labelling expert
- Help UN experts sort out Voluminous SDG articles into their various categories

● Goal

- Our team aims to develop a quick and objective way to analyze large quantities of UN publications according to a desired criteria using Natural Language Processing model

● Results

- Built a streamlit web-based app capable of classifying articles into SDGs





DATA CLEANING/PREPROCESSING

● Data Integrity Cycle



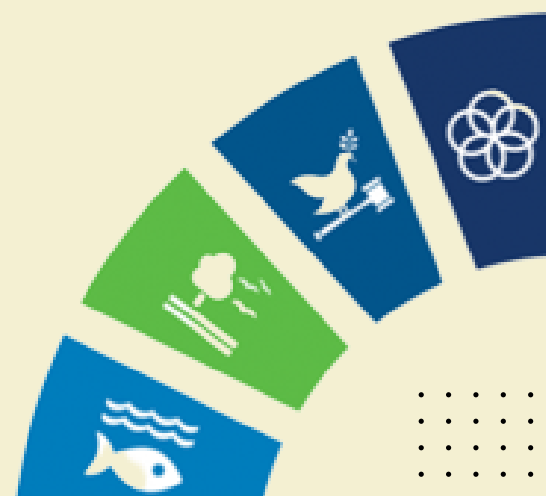
● Data preparation for model

● Sorting out articles

- Agreement score > 0.94
- Cumulative Probability > 0.58

● Natural Language Processing

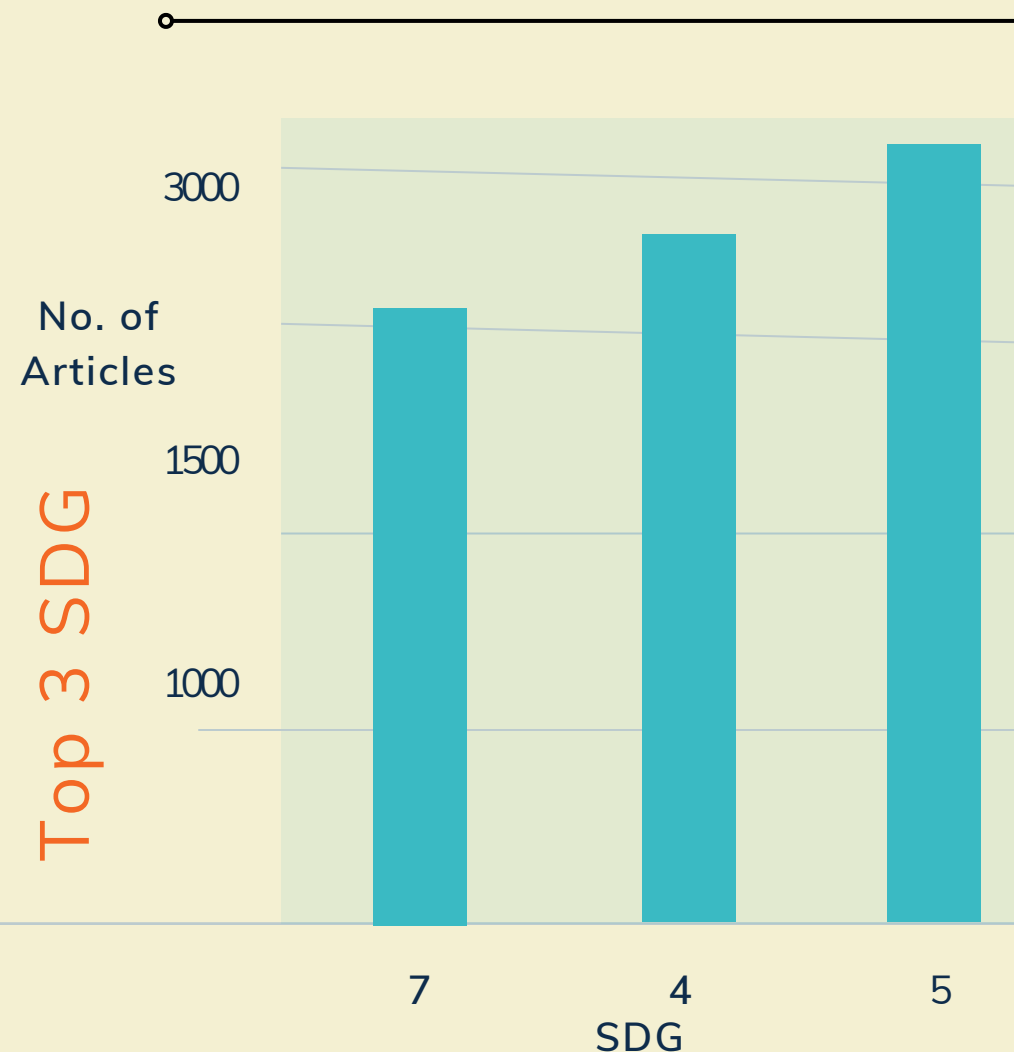
- Remove noises
- Removed Stopwords
- Tokenization
- Lemmatization





DATA ANALYSIS/VISUALIZATION

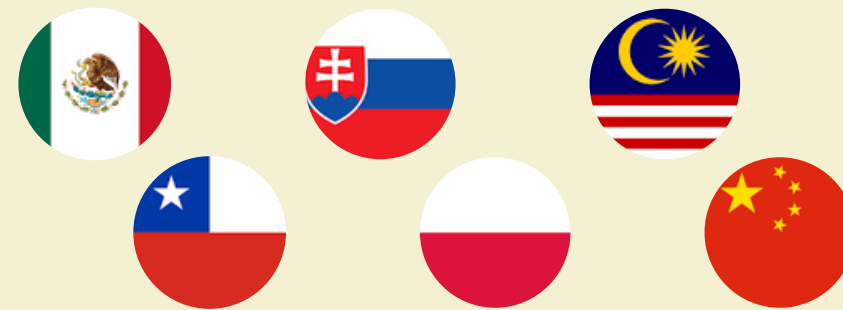
• Most Frequent SDGs



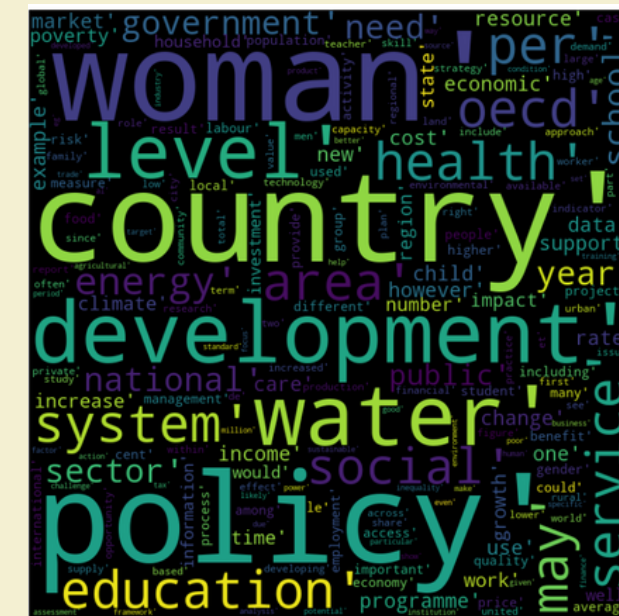
- Gender equality (5)
- Quality Education (4)
- Affordable and clean energy (7)

Countries Involved in top 3 SDGs

- SDG 5: Mexico & Slovakia
- SDG 4: Malaysia & Chile
- SDG 7: Poland & China

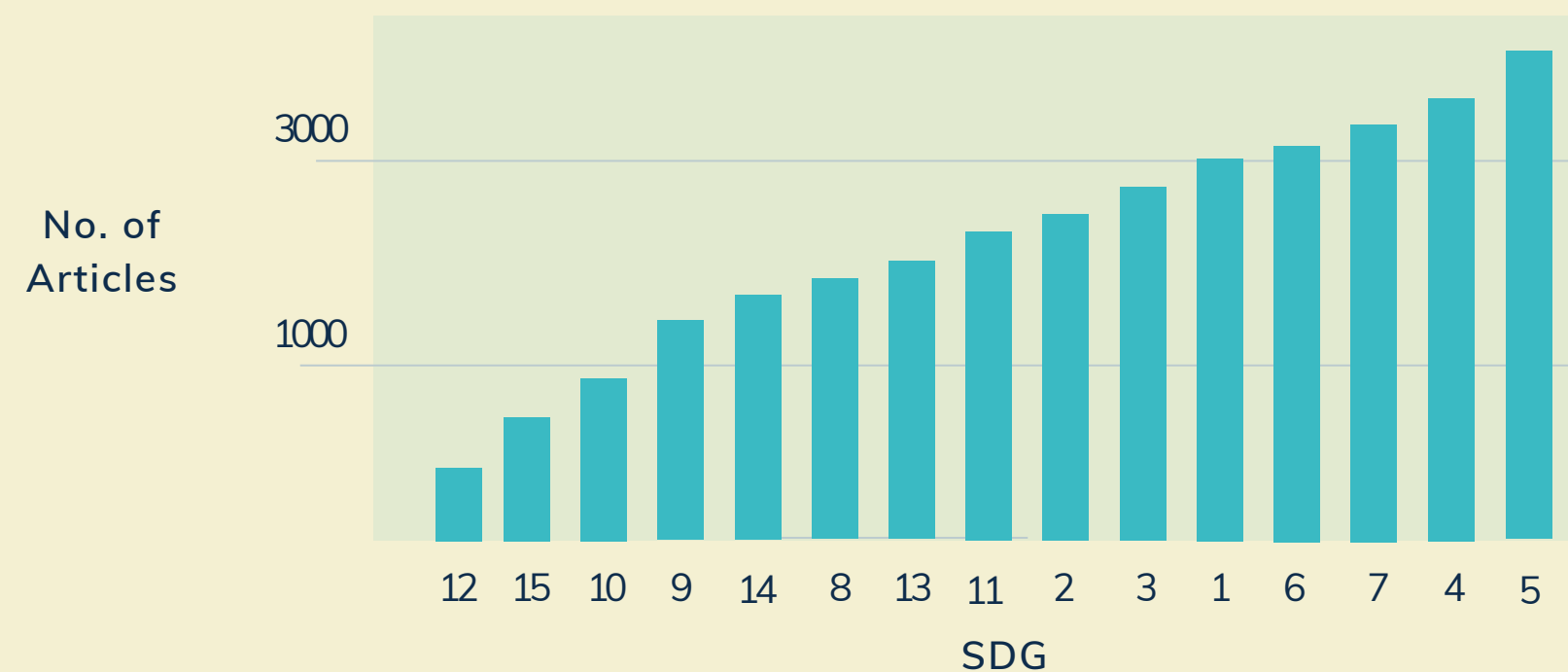


Text Analysis (Most frequent words)



- Country
- Policy
- Women
- Development
- Water

SDG Distribution of data by No. of articles





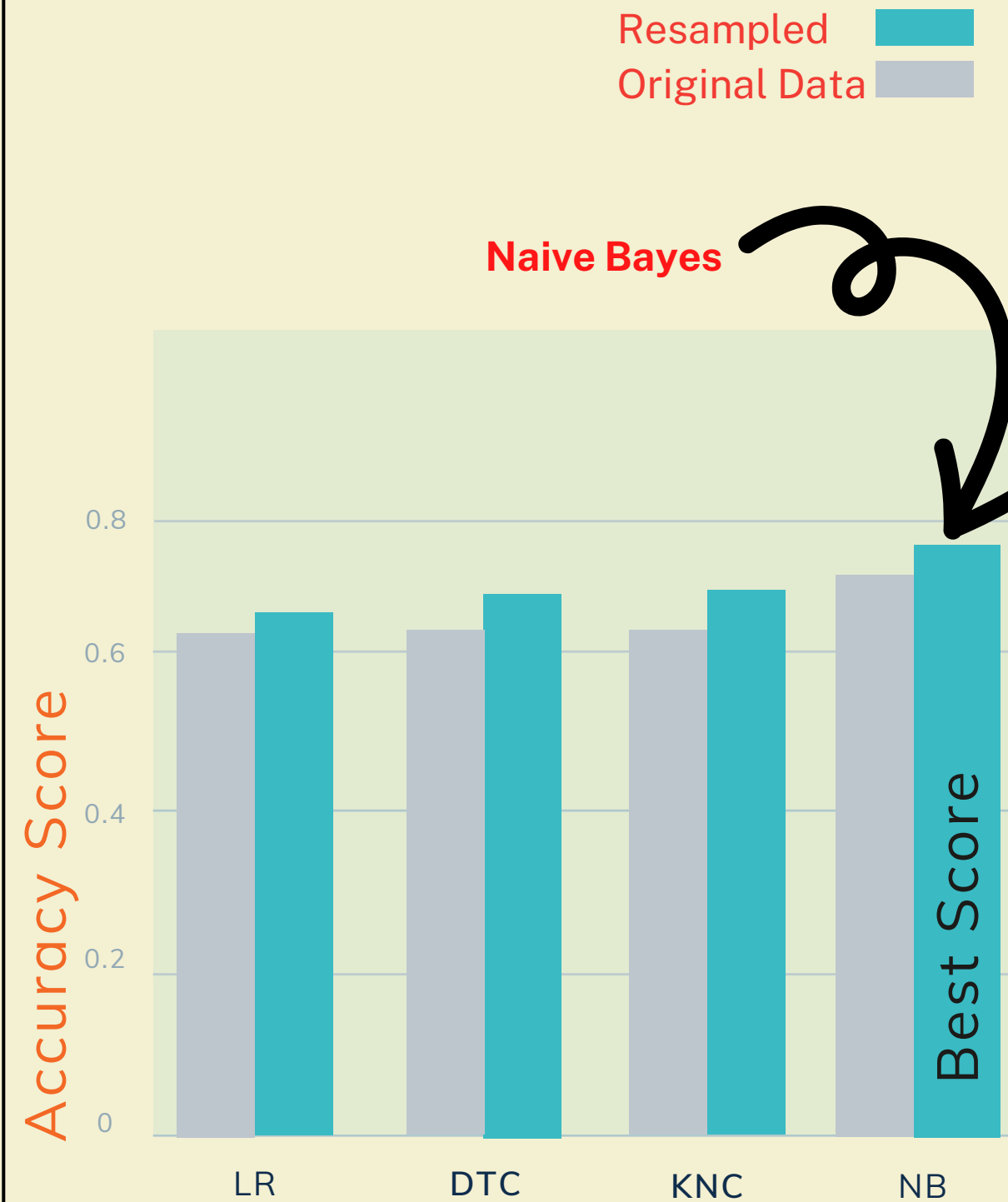
MODEL

WHY NAIVE BAYES MODEL?

- Simple & easy to implement
- Highly scalable
- Real-time predictions
- High speed with big data

HOW THE MODEL WORKS

- Calculate
- Find
- Input
- Check



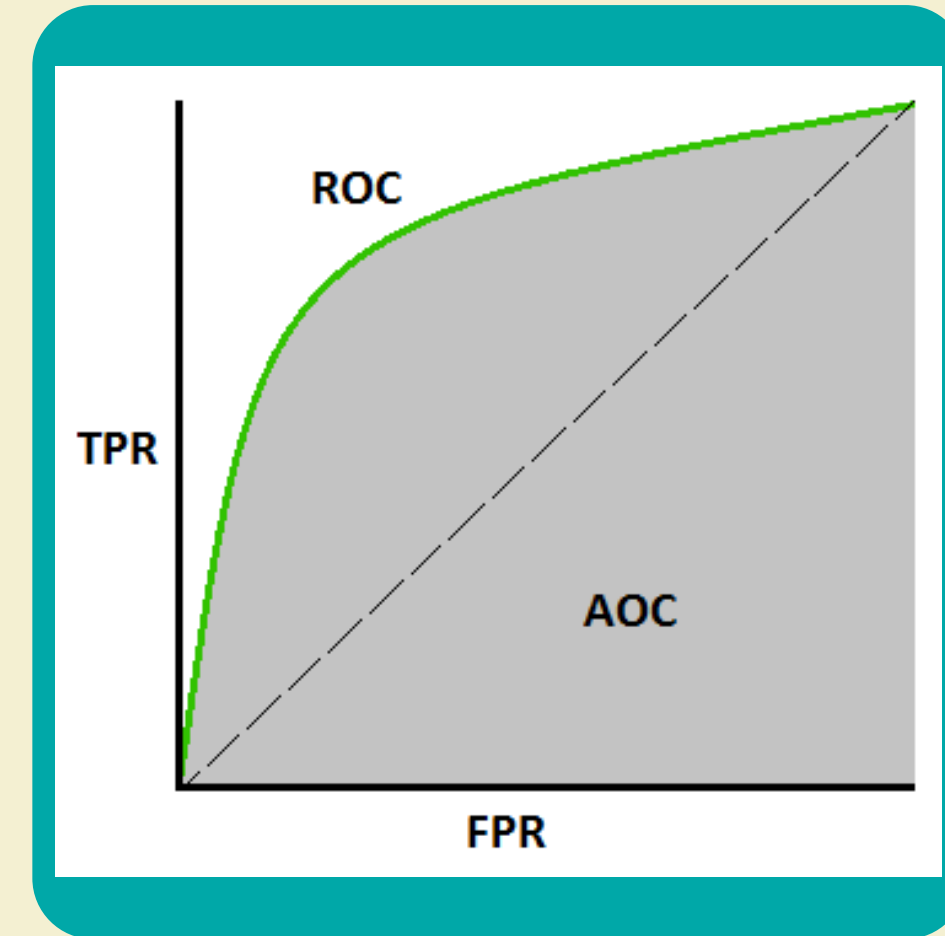
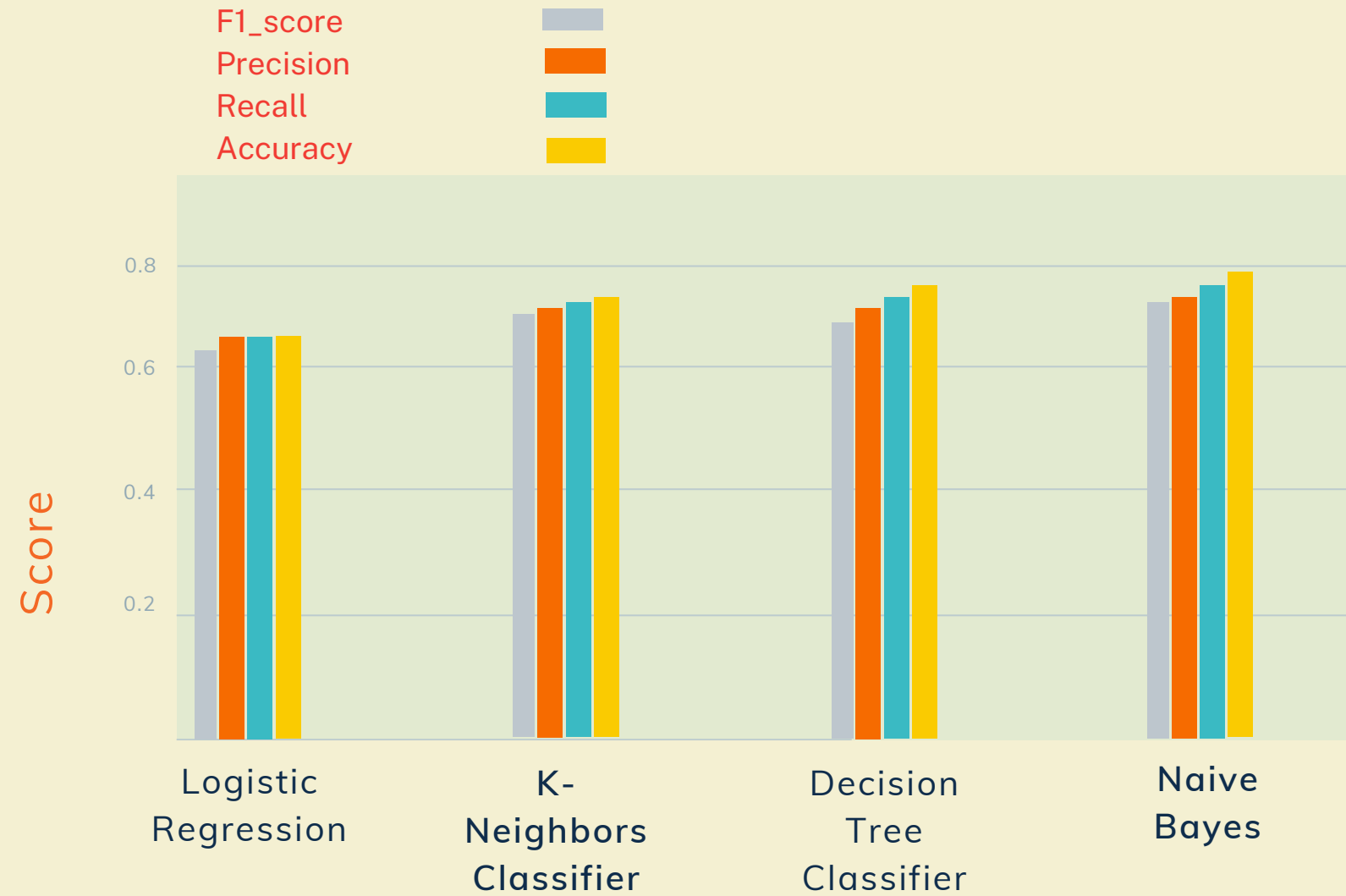
OTHER MODELS USED

- Logistic Regression
- Decision Tree Classifier
- K- Neighbour Classifier
- Bert



EVALUATION METRICS

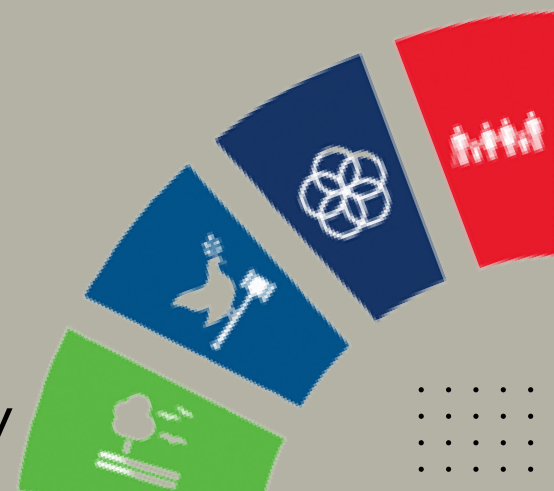
MODELS UTILISED



Suggested metric: Area Under The Curve - Receiver Operating Characteristics curve (AUC - ROC)

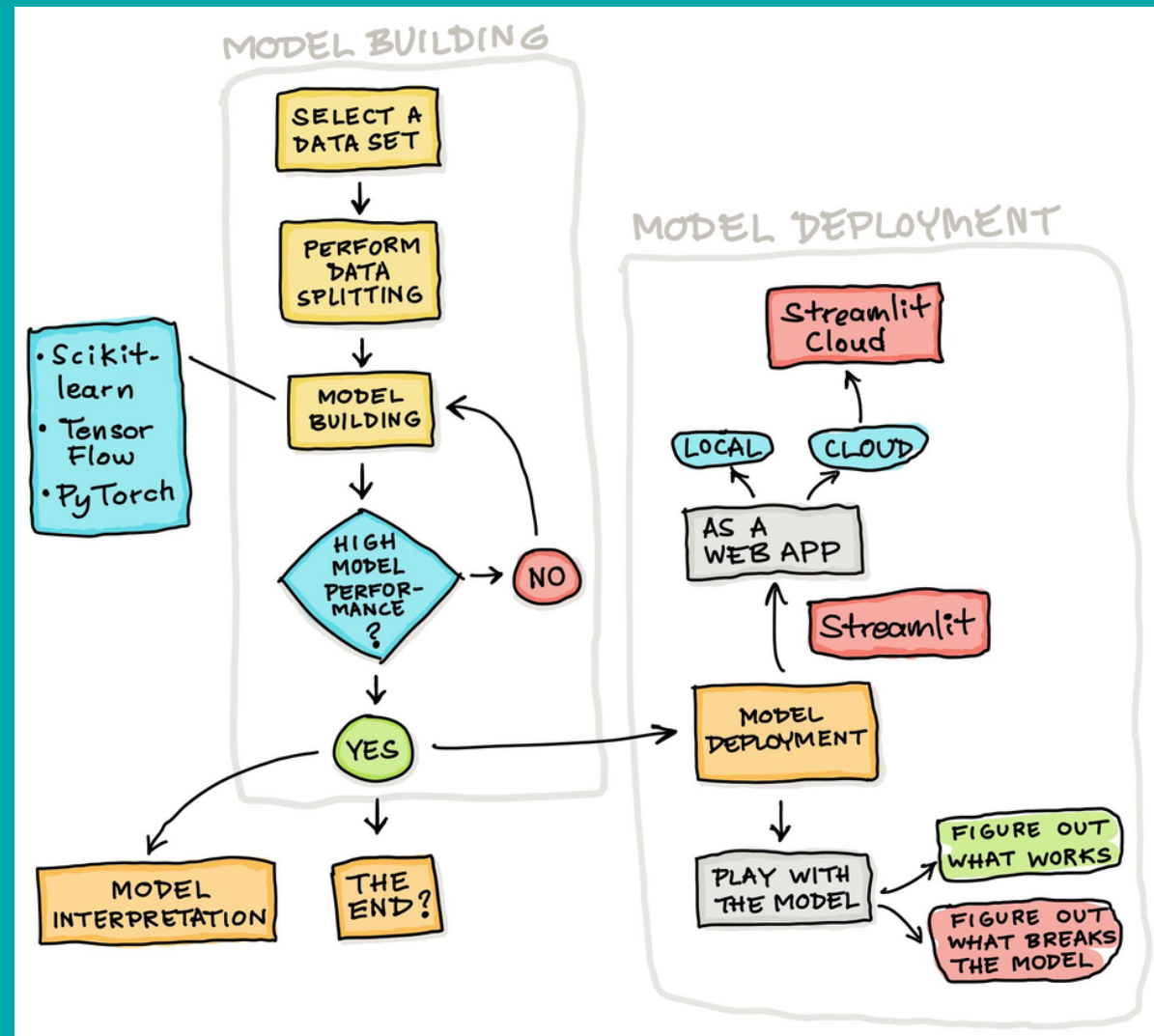
Metric evaluation
if AUC is
near 1 - good
near 0 - bad

0.5 - no class separation capacity





DEPLOYMENT



Architecture

Short-Term Recommendation

- The fastest way to deploy with few lines of code. (Stream lit)

Long-Term Recommendation

- Docker

Alternative Deployment Tools

- SageMaker
- Flask

Use of the App in real world

- Classify CSR reports
- Labeling SDGs



TEAM RETROSPECTIVE



What went well?

- Successful Team Collaboration
- Effective Communication
- Team Availability and Support

Pain points?

- Model Improvement
- Project standard was below expectation

What to do differently?

- Do more Technical research
- Establish timelines before commencement of project



APPENDIX

Deployment

- <https://yussufgit-deployment-main-6duok6.streamlitapp.com/>