



Predicting National Competitiveness Based on Social Economic, and Environmental Indicators

Prasetyo Widyantoro
Continuation Batch 8

Pembahasan

01

Introduction

02

Dataset and Features

03

Metode

04

Analisa dan kesimpulan

05

Future Works



Introduction

Melalui pendekatan globalisasi, perkembangan dunia tanpa disadari telah semakin bergerak cepat, berbagai negara mendapati sejumlah tantangan yang menghingapi, salah satu yang terbesar yang mesti ditanggapi ialah perihal persaingan antar negara.

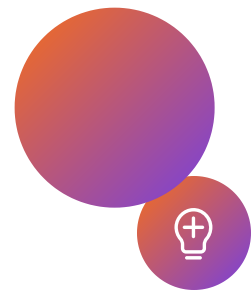
Saat ini, daya saing nasional merupakan suatu identitas tingkat keberhasilan negara dalam mengelola sumber daya yang dimiliki untuk mempertahankan tingkat produktivitas, pada akhirnya mendorong pembangunan berkelanjutan yang menentukan tingkat kesejahteraan warga negara (Schwab, 2018).

Problem

Saat ini pengukuran daya saing nasional dikeluarkan oleh World Economics Forum dalam bentuk Global Competitiveness Report setiap tahunnya. Namun belakangan ini report tersebut tidak dikuarkan oleh lembaga yang bersangkutan. Hal ini memerlukan alternative agar setiap negara mengetahui capaian daya saing negaranya

Solution

Memprediksi nilai daya saing nasional dengan acuan report tahun terakhir terbit yakni pada tahun 2019 menggunakan data indikator sosial ekonomi dan lingkungan dari setiap negara dan diproses dengan menggunakan analisis regresi.



Dataset & Feature

Sumber data yakni Bank Dunia, Forum Ekonomi Dunia, Global Competitiveness Report, UNDP, dan macrotrends pada tahun 2019, dengan jumlah data sebanyak 13 kolom dan 141 baris

Nama Kolom	Keterangan	Tipe Data
Country_name	Sampel negara yang diobservasi	String
gci	Global Competitiveness Index (1-100)	Numerikal
gdpp	Gross Domestic Product Percapita (\$)	Numerikal
fdi	Foreign Direct Investment (\$)	Numerikal
inflation	Inflasi tahunan negara	String (akan dilakukan penggantian menjadi numerik)
healthcare_spend	pengeluaran kesehatan setiap warga negara (\$)	Numerikal
Unemployment_rate	tingkat pengangguran dalam persen	Numerikal



Dataset & Feature

Nama Kolom	Keterangan	Tipe Data
carbon_emissions	jumlah emisi karbon dalam metrics ton	Numerikal
Export	Nilai ekspor perkapita (\$)	Numerikal
Imports	Nilai import percapita (\$)	Numerikal
eyys	Ekspektasi lama sekolah	Numerikal
mys	Rata-rata lama sekolah	Numerikal
gender_inequality	gender inequality index (0-1)	Numerikal



Feature yang digunakan untuk Model

Variabel Dependend (y)

- gci

Variabel Independend (x)

- gdpp
- fdi
- inflation
- healthcare_spend
- Unemployment_rate
- carbon_emissions
- Export
- Imports
- mys
- gender_inequality

Metode

Analysis Regresi dengan goal meminimalkan nilai MAE

$$\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

n: mewakili jumlah pengamatan
y_i: mewakili nilai Aktual
ŷ_i: mewakili nilai Prediksi

Memprediksi nilai output sejumlah satu kolom yang berdasarkan kalkulasi dari nilai input sejumlah 10 kolom

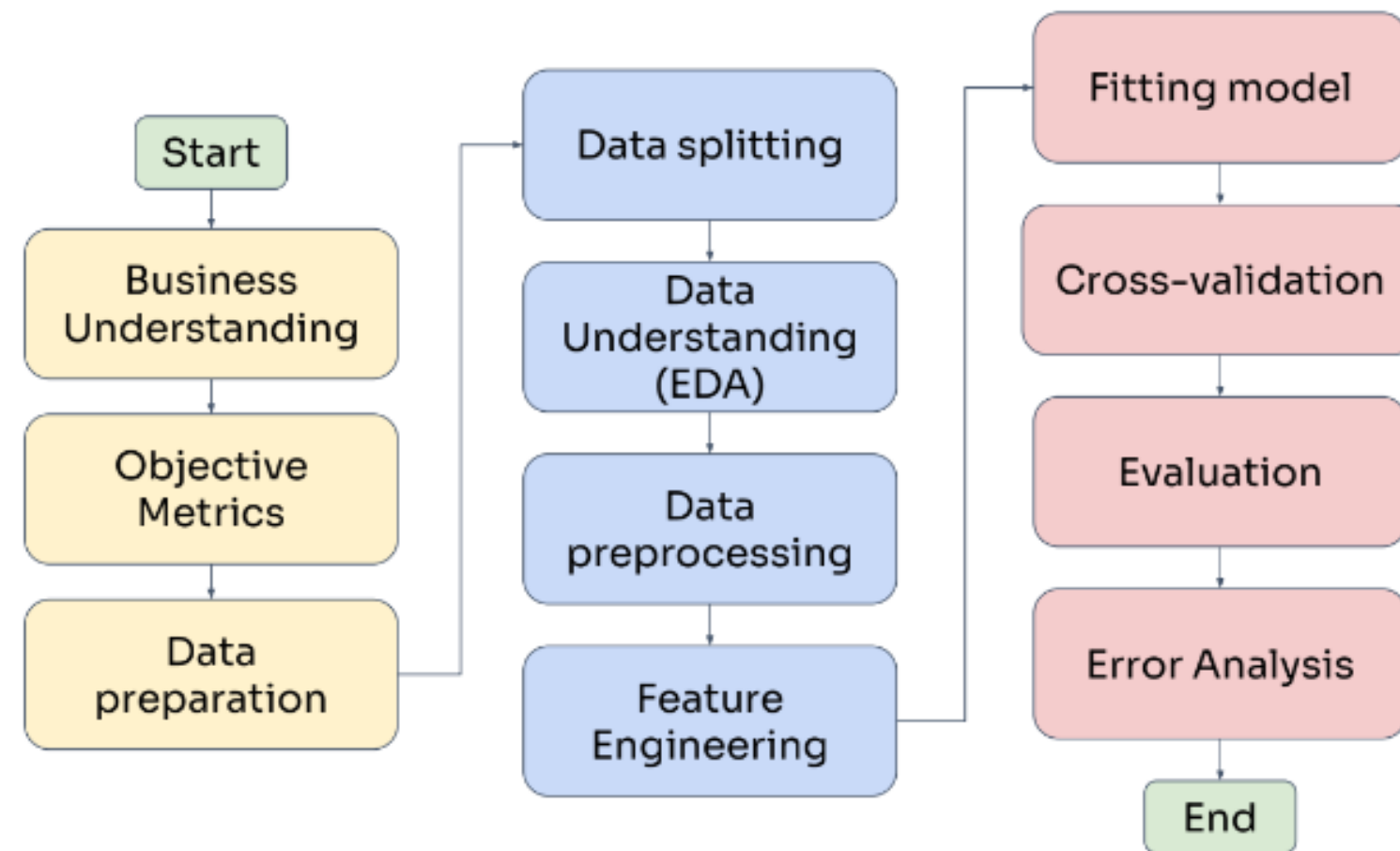
$$f : \mathbb{R}^{10} \rightarrow \mathbb{R}^1$$

Metode Regresi Analysis yang digunakan untuk eksperimentasi pemodelan

- decisiontree
- extratrees
- linear regression
- randomforest regression
- bagging
- adaboost
- gradientboost
- XGboost
- KNN Regression



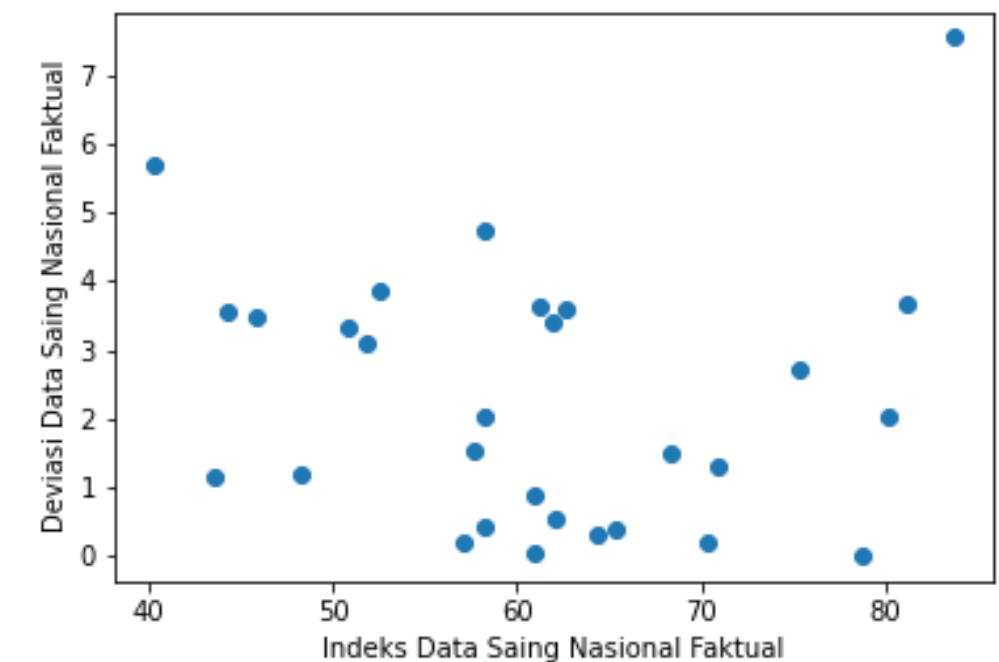
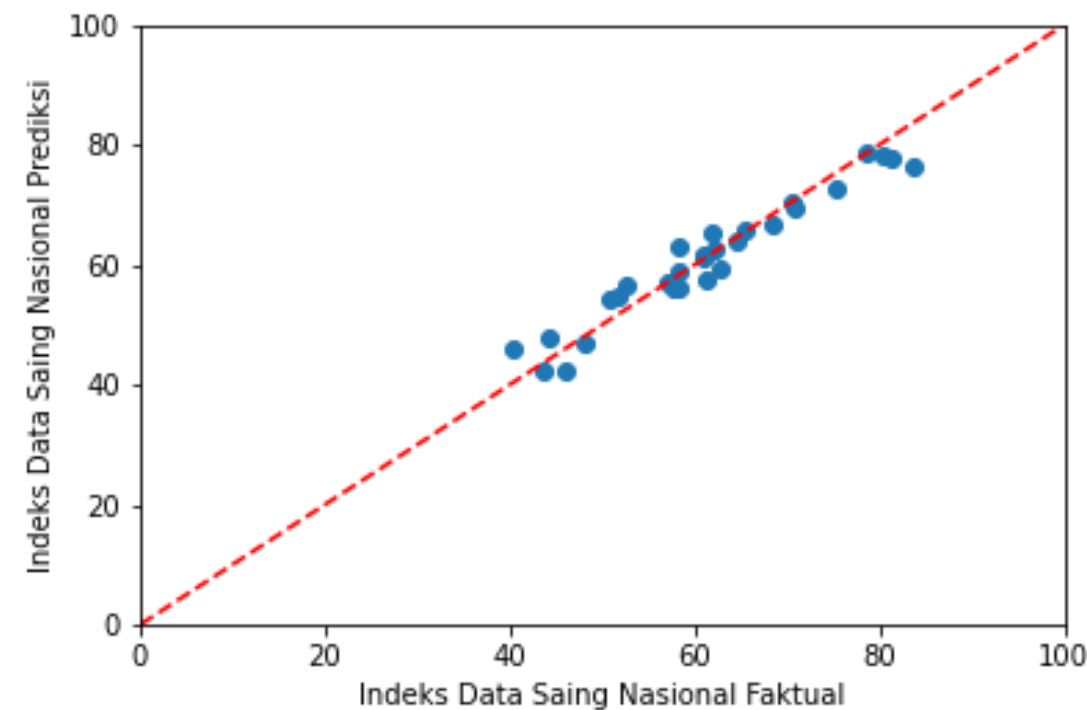
Proses Pemodelan





Hasil Pemodelan

	MAE Train	MAE CV
lin_reg	24.471752	-
decision tree	3.024946	4.574225
ExtraTrees	0.156871	3.466897
bagging	1.285491	3.61731
random forest	1.195786	3.423697
adaBoost	2.187123	3.990832
gradientBoost	0.876317	3.574963
XGboost	0.000504	3.528157
Knn	3.504847	0.823745



Jadi, penulis memutuskan untuk menggunakan randomforest sebagai model yang digunakan untuk memprediksi daya saing nasional setiap negara

Future Works

- Menggunakan teknik ekonometrika untuk mengetahui pengaruh antar variabel
- Mencoba model regresi lain berbasiskan neural network maupun lightGBM
- Dapat menggunakan PCA dan dilanjutkan dengan analisis regresi



Terimakasih.