

EE_Predict: Machine Learning Pemrediksi Produksi Listrik dan Emisi Karbon Indonesia

Disusun Oleh:

Revyra Phoebe Fransiska / 105217020 / rephfr@gmail.com
Tiurmauli Rahel Ernita / 105217029/ acheellernita@gmail.com



Laporan penelitian ini adalah sebagai bentuk
Ujian Akhir Semester (UAS)
untuk mata kuliah Mesin Pembelajaran (CS-0037)

Mei 2019

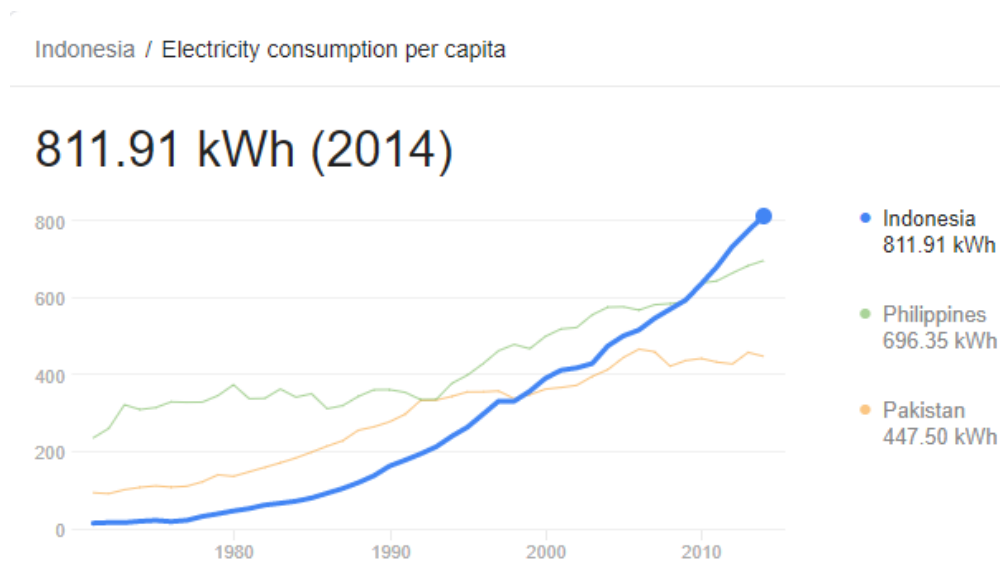
DAFTAR ISI

I.	PENDAHULUAN	3
1.	Latar Belakang.....	3
2.	Manfaat	4
II.	RANCANGAN SISTEM YANG DIUSULKAN	4
III.	EVALUASI DAN ANALISIS.....	4
IV.	KESIMPULAN	6
V.	REFERENSI.....	11
	LAMPIRAN	11

I. PENDAHULUAN (10 points)

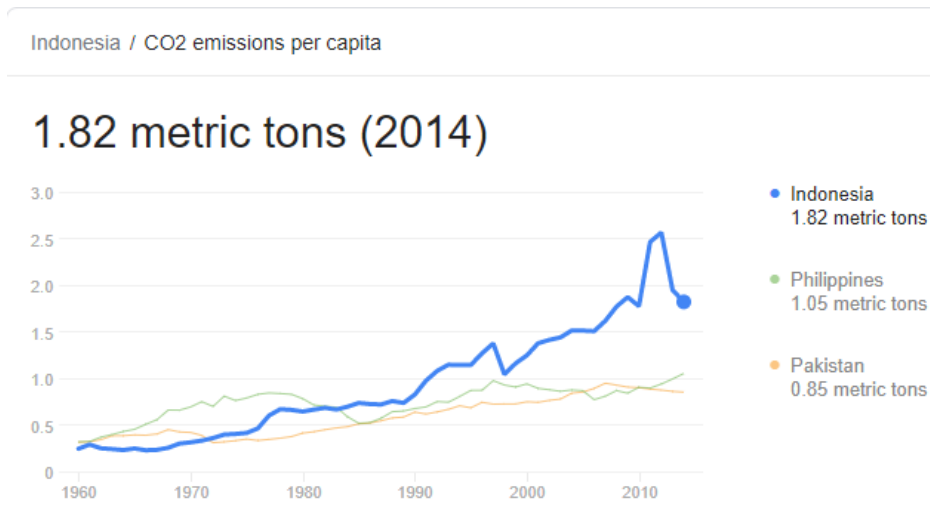
1. Latar Belakang

Energi listrik merupakan salah satu energi yang penting bagi masyarakat karena aktivitas - aktivitas manusia yang banyak melibatkan energi listrik, contohnya seperti penggunaan alat - alat elektronik dan penggunaan lampu sebagai penerangan di rumah. Seiring perkembangan zaman, produksi dan konsumsi energi listrik negara Indonesia terus meningkat. Hal ini menyebabkan emisi karbon yang sangat tinggi, Indonesia menduduki posisi ke-10 penghasil emisi karbon terbesar di dunia pada tahun 2016.



Gambar 1.2 Emisi karbon di Indonesia hingga tahun 2014

Emisi karbon memiliki dampak yang sangat buruk terhadap lingkungan, seperti pemanasan global, cuaca tak tentu, perubahan iklim, dan sebagainya. Salah satu upaya yang dapat dilakukan untuk menurunkan emisi karbon adalah dengan setidaknya menyeimbangkan produksi listrik antara sumber Energi Baru dan Terbarukan (EBT) dan non EBT. EBT menghasilkan sedikit emisi karbon hingga tidak sama sekali. Sejauh ini meski Indonesia sudah mulai menggunakan EBT untuk memproduksi listrik, persentase produksi listrik dari energi non EBT masih sangat tinggi dibandingkan EBT.



Gambar 1.2 Emisi karbon di Indonesia hingga tahun 2014

Oleh sebab itu kami membuat rancangan program *machine learning* bernama “EE_Predict” menggunakan algoritma *supervised learning* berjenis regresi linear dan *k neighbor regressor* dari Scikit-Learn untuk memprediksi seberapa besar emisi karbon yang dihasilkan berdasarkan beberapa faktor seperti populasi, konsumsi listrik, perbandingan produksi energi listrik EBT dan non EBT, dsb. Kemudian kami juga akan memprediksi produksi listrik kedua sumber utama ini untuk menurunkan emisi karbon.

2. Manfaat

Dengan EE_Predict, kami dapat mengetahui seberapa besar produksi listrik dari EBT dan non EBT beserta emisi karbon yang dihasilkannya, data prediksi ini dapat digunakan pemerintah misalkan kementerian ESDM dan industri pembangkit tenaga listrik Indonesia untuk dapat memperkirakan seberapa besar rasio produksi listrik dari EBT dan non EBT yang pas untuk mengurangi emisi karbon. Berkurangnya emisi karbon yang dihasilkan Indonesia akan berkontribusi terhadap kelestarian lingkungan, tidak hanya bagi negara ini saja tetapi bagi dunia.

II. RANCANGAN SISTEM YANG DIUSULKAN (10 points)

Program EE_Predict ini dirancang agar dapat mengolah *input* populasi, produksi listrik, total emisi karbon, konsumsi energi, yang kemudian akan menghasilkan *output* prediksi emisi karbon dari listrik. Oleh karena itu program ini membutuhkan *dataset* yang berisikan data - data milik negara Indonesia per tahunnya, berikut data - data yang ada pada *dataset* ini, antara lain:

- Tahun
- Data total produksi listrik (kWh)
- Data produksi listrik dari sumber EBT (kWh)
- Data produksi listrik dari sumber non EBT (kWh)
- Data konsumsi listrik (kWh)
- Data populasi
- Data total emisi karbon (kt)

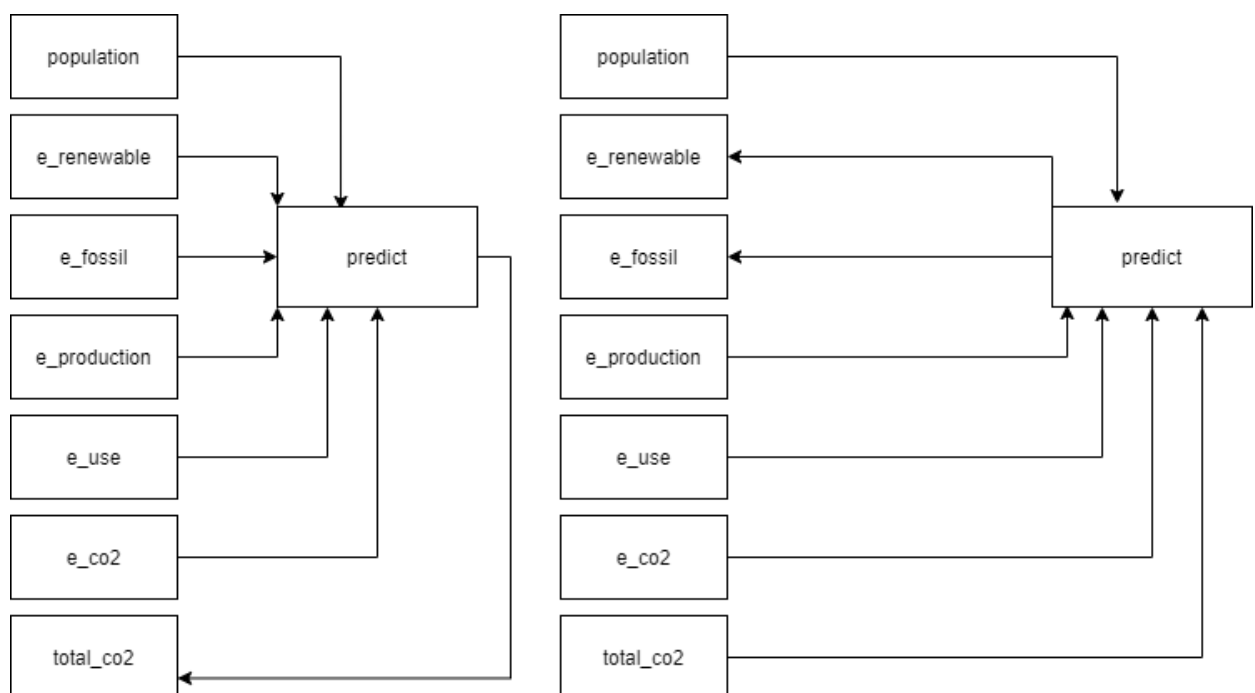
- Data emisi karbon dari listrik (kt)
- Sumber
 - <https://data.worldbank.org/indicator/SP.POP.TOTL?locations=ID>
 - <https://data.worldbank.org/country/indonesia>
 - <https://ourworldindata.org/grapher/renewables-share-electricity-production>

Data - data diatas diambil dari sumber *website* The World Bank dan Our World in Data, kami hanya mengambil beberapa data penting saja yang berhubungan dengan studi kasus kami dan memasukkannya ke dataset “energy_impact_upd”. Dataset ini memiliki merupakan *file-based dataset* dengan format csv dan berukuran (59, 8).

Strategi kami dalam penggunaan *dataset* pada program ini adalah dengan membagi data *training* dan *test*, dengan data *test* sebesar 20% dan data *training* sebesar 80%.

Library yang kami gunakan antara lain numpy, pandas, dan sklearn. Numpy dan pandas digunakan untuk pemrosesan data yang mendasar, misal untuk melakukan perhitungan matematis dan untuk membaca maupun mengolah *dataset*. Di dalam kami memakai *train_test_split* untuk membagi data untuk *training* dan untuk *testing*, kemudian *LinearRegression* untuk melakukan regresi linear agar dapat memprediksi data menggunakan model linier, lalu *KNeighborsRegressor* kami gunakan untuk melakukan prediksi dengan model tetangga terdekat.

Rencananya kami akan membuat dua prediktor, yang satu akan memberikan prediksi emisi karbon dari listrik dan yang satu akan memprediksi produksi listrik dari sumber EBT dan non EBT. Berikut ini adalah *block diagram*nya.



Gambar 2.1 Block Diagram EE_Predict

III. EVALUASI DAN ANALISIS (15 points)

```
[129]: import pandas as pd
import numpy as np

[4]: data = pd.read_csv("energy_impact_upd.csv")
data.head()
```

Gambar 3.1 Pemanggilan *library* dan pembacaan *dataset*.

Pada Gambar 3.1, ditunjukkan bahwa program EE_Predict ini melakukan *import library* pandas dan numpy. Kemudian, program ini akan membaca *dataset* "energy_impact_upd" sebagai acuan program ini untuk melakukan prediksi emisi karbon dan produksi listrik di Indonesia.

```
data.columns = ['year', 'e_production', 'e_use', 'population', 'total_co2', 'e_fossil', 'e_renewable', 'e_co2']
data.shape

(59, 8)

data_tanpa_NaN = data[(data["year"]>=1982) & (data["year"]<=2013)]

x = data_tanpa_NaN.loc[:, ['e_production', 'e_use', 'population', 'total_co2', 'e_fossil', 'e_renewable']]
y = data_tanpa_NaN["e_co2"] #data emisi karbon dari listrik
#X = data_tanpa_NaN.iloc[:, :-1].values
#Y = data_tanpa_NaN.iloc[:, 7].values
#X=X[:, 1:]

data_tanpa_NaN.head()
```

Gambar 3.2 Mengubah nama kolom serta membagi data x dan y untuk prediksi emisi karbon

Pada bagian ini kami menamai ulang setiap kolom yang ada pada dataset untuk mempermudah proses pengolahan data. Lalu, kami menentukan range dari tahun 1982 sampai tahun 2013, sebab beberapa data yang ada pada dataset ini masih tidak lengkap (bernilai NaN), jika ini tidak dilakukan nantinya program akan error ketika ingin melakukan prediksi. Kemudian, program ini membagi 2 data menjadi data x dan y, dimana y adalah data emisi karbon dari produksi listrik dan x merupakan data semua kolom kecuali x.

```
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(x, y, test_size = 0.2, random_state = 0)
```

Gambar 3.3 pembagian data untuk *training* dan *testing* untuk prediksi emisi karbon

Selain itu, program ini menggunakan *library* sklearn untuk melakukan *splitting*, *training*, dan *testing* data. Untuk melakukan *training* dan *testing*, program ini membagi datanya untuk memprediksi emisi karbon. Jadi ada data untuk *training* dan ada data untuk *testing*. Data *training* ini digunakan untuk melakukan fitting data, sedangkan *testing* digunakan untuk menguji atau mendeteksi kebenaran pada model tersebut.

```
from sklearn.linear_model import LinearRegression
linRegressor = LinearRegression()
linRegressor.fit(X_train, Y_train)
```

Gambar 3.4 Import *library* regresi linier dan melakukan fitting data untuk prediksi emisi karbon

Prediksi emisi karbon program ini menggunakan konsep regresi linier. Oleh karena itu, program ini melakukan *import library* LinearRegression dari sklearn. Lalu data yang sudah dibagi sebelumnya (data milik *training*) pada program ini akan di *fitting* untuk melakukan *training* data.

```
Y_predCO2_1 = linRegressor.predict(X_test)

print(linRegressor.score(X_test, Y_test)*100, "%")
#akurasi test

98.1582974927268 %

Y_predCO2_1

array([ 79158.02392993, 132362.88794583,  71596.84226201, 27023.00367716,
        73713.92009018,  94479.37168653, 174792.21318207])
```

Gambar 3.5 Melakukan prediksi dengan regresi linear dan kalkulasi koefisien determinasi untuk prediksi emisi karbon

Setelah itu, program ini akan memprediksi emisi karbon dengan model regresi linear dan menghitung koefisien determinasinya yang digunakan sebagai nilai akurasi tes. Koefisien ini menggambarkan kualitas *output* yang direplikasi oleh model. Apabila nilai koefisien determinasi semakin tinggi, itu artinya tes yang dilakukan mempunyai akurasi yang lebih baik. Pada Gambar 3.5 ini akan menampilkan hasil prediksi emisi karbon.

```
from sklearn.neighbors import KNeighborsRegressor
neigh = KNeighborsRegressor(n_neighbors=1, metric='euclidean')

neigh.fit(X_train, Y_train)

KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='euclidean',
                    metric_params=None, n_jobs=None, n_neighbors=1, p=2,
                    weights='uniform')

Y_predCO2_2 = neigh.predict(X_test)
print(neigh.score(X_test, Y_test)*100, "%")
Y_predCO2_2

95.50538253472213 %

array([ 79157.66636, 124631.4509,  67159.89897, 30018.75003,
        104647.4794,  81800.77543, 180780.1621 ])
```

```

predData1 = neigh.predict(pd.DataFrame(
    [[169755000000, 680.2814172, 241834215, 428760.308, 142841999997.054, 26913000123.5001]]))
print(predData1) #data 2010
trueData1 = 172915.305102906

180780.1621

print(abs(trueData1-predData1)/trueData1*100,"%")

4.548386848933609 %

```

Gambar 3.6 Model kNN, *training*, prediksi, dan kalkulasi koefisien determinasi model regresi *nearest neighbor* untuk prediksi emisi karbon

Selain model regresi linear, program ini juga menggunakan model kNN untuk memprediksi emisi karbon. Tujuan dari penggunaan model regresi linear dan kNN adalah untuk memiliki dua pendekatan berbeda, karena setiap model yang digunakan pasti mempunyai nilai akurasi yang berbeda. Pada model kNN ini, program ini melakukan hal yang sama seperti model regresi linier yaitu sama - sama melakukan fitting data, training data, prediksi, dan menghitung koefisien determinasi dari model ini.

Saat menggunakan model ini, nilai akurasi yang dihasilkan lebih kecil daripada nilai akurasi dari model regresi linear. Hal ini dapat memberi pernyataan bahwa model regresi linear lebih baik daripada model kNN untuk kasus prediksi jenis ini. Hal ini dapat disebabkan oleh berapa jumlah *neighbour* yang digunakan, berdasarkan beberapa kali *trial and error*, kami dapati bahwa untuk memprediksi emisi karbon, jumlah *neighbour* yang tinggi akan menurunkan akurasi atau koefisien determinasi, maka dari itu jumlah *neighbour* yg kami gunakan adalah satu. Selain dari itu kNN dapat dikatakan sebagai *lazy learning algorithm*, ini juga akan sangat bermain pada akurasi data, sebab kNN hanya belajar dengan mengingat data *training*. Algoritma ini juga merupakan *non-parametric model*, sehingga tidak dapat dikarakterisasi oleh sekumpulan parameter tetap dan tidak ada jenis distribusi statistika yang benar-benar cocok untuk setiap data set, seberapa baik jumlah *neighbour*-nya pun sangat bergantung pada masing-masing *dataset*.

Setelah itu, program ini akan memberikan output nilai prediksi emisi karbon dengan satuan kt dan dibandingkan dengan data sebenarnya. Pengguna dapat menginput data mereka sendiri untuk dapat mengetahui seberapa besar emisi karbon yang akan dihasilkan berdasarkan data populasi, produksi listrik, dsb. Ketika kami menggunakan data tahun 2010, error relatif prediksi ini adalah 4.4584%.

```

x2 = data_tanpa_NaN.loc[:,['e_production','e_use','population', 'total_co2', 'e_co2']]
y2 = data_tanpa_NaN.loc[:,['e_fossil', 'e_renewable']]
#X2 = x2.values
#Y2 = y2.values
#X2=X2[:, 1:]

```

Gambar 3.7 Pembagian data x dan y untuk prediksi produksi listrik

Begitu juga dengan melakukan prediksi produksi listrik, langkah pertama yang harus dilakukan adalah membagi lagi data x dan y, dengan y adalah data produksi listrik sumber EBT dan non EBT, sedangkan x adalah data semua kolom kecuali y.

```
X2_train, X2_test, Y2_train, Y2_test = train_test_split(x2, y2, test_size = 0.2, random_state = 0)
linRegressor2 = LinearRegression()
linRegressor2.fit(X2_train, Y2_train)

LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

Gambar 3.8 Pembagian data *training* dan *testing* serta melakukan *fitting data* untuk prediksi produksi listrik (regresi linier)

Selanjutnya membagi data menjadi dua bagian. Satu untuk digunakan pada saat melakukan *training*, yang satunya digunakan pada saat melakukan *testing*. Pada kali ini, program melakukan *training*, *testing*, dan *fitting data* dengan model regresi linier untuk memprediksi produksi listrik.

```
Y_predListrik = linRegressor2.predict(X2_test)
print(linRegressor2.score(X2_test, Y2_test)*100, "%")
99.42692989401972 %

Y_predListrik

array([[3.75243440e+10, 7.92135854e+09],
       [1.03544015e+11, 1.64695742e+10],
       [3.39177134e+10, 7.65526958e+09],
       [1.03491014e+10, 3.27931306e+09],
       [6.53757227e+10, 1.26267023e+10],
       [5.77639109e+10, 9.95071939e+09],
       [1.48336944e+11, 2.16486161e+10]])
```

Gambar 3.9 Prediksi produksi listrik sumber EBT dan non EBT serta pengkalkulasian koefisien determinasinya (regresi linier)

Kemudian, program ini melakukan prediksi listrik sumber EBT dan non EBT dan menghitung koefisien determinasinya untuk mengetahui nilai akurasi saat menggunakan model regresi linier.

```
neigh2 = KNeighborsRegressor(n_neighbors=1, metric='euclidean')

neigh2.fit(X2_train, Y2_train)

KNeighborsRegressor(algorithm='auto', leaf_size=30, metric='euclidean',
                    metric_params=None, n_jobs=None, n_neighbors=1, p=2,
                    weights='uniform')

Y2_predListrik = neigh2.predict(X2_test)
print(neigh2.score(X2_test, Y2_test)*100, "%")
Y2_predListrik
97.59966060166569 %
```

```

array([[4.33170000e+10, 8.04100023e+09],
       [9.75660000e+10, 1.54080003e+10],
       [2.96640000e+10, 7.67600009e+09],
       [1.08840000e+10, 3.21400002e+09],
       [6.69110000e+10, 7.71600022e+09],
       [6.69110000e+10, 7.71600022e+09],
       [1.36028000e+11, 2.07459993e+10]])

predData2 = neigh2.predict(pd.DataFrame([[169755000000,680.2814172,241834215,428760.308,172915.305102906]]))[0]
print(predData2)
#data tahun 2010
trueData2 = [142841999997.054, 26913000123.5001]

[1.36028000e+11 2.07459993e+10]

print(abs(trueData2-predData2)/trueData2*100,"%")

[ 4.77030565 22.91457965] %

```

Gambar 3.10 Model kNN, *training*, prediksi, dan kalkulasi koefisien determinasi model regresi *nearest neighbor* untuk prediksi produksi listrik

Setelah itu, program ini juga melakukan *training*, *fitting* data, dan prediksi dengan menggunakan model kNN. Lalu, menghitung kembali koefisien determinasi sebagai nilai akurasi test dengan menggunakan model kNN. Perbandingan dari nilai akurasi model regresi linear dan model kNN adalah model regresi linear mempunyai nilai akurasi yang lebih besar daripada model kNN. Hal ini juga membuktikan bahwa model regresi linear lebih baik daripada model kNN untuk kasus prediksi jenis ini dengan alasan yang sama seperti pada bagian prediksi emisi karbon. Kemudian, program ini akan memberikan *output* prediksi nilai produksi listrik sumber non EBT dan EBT dengan satuan kWH berdasarkan data-data seperti populasi, emisi, dsb.

Pengguna dapat menginput data mereka sendiri, misalkan mereka ingin memprediksi ketika populasi di Indonesia mencapai sekian, kemudian berapa emisi karbon yang dikehendaki, dan data-data lainnya. Program akan memberikan prediksi berapa produksi listrik sumber non EBT dan EBT, ini dapat digunakan bagi pemerintah atau pembangkit tenaga listrik untuk membuat rencana seberapa besar perbandingan produksi per sumbernya untuk menurunkan emisi karbon.

Dari gambar di atas dapat dilihat bahwa eror relatif prediksi produksi listrik non EBT adalah 4.7703% sedangkan untuk produksi listrik EBT sebesar 22.9146%.

IV. KESIMPULAN (5 points)

Berdasarkan pekerjaan kami di atas, dapat disimpulkan bahwa program *machine learning* EE_Predict ini akan membantu Indonesia mengatasi dampak dari emisi karbon dengan memprediksi seberapa banyak produksi listrik yang dihasilkan per sumbernya atau emisi karbon yang dihasilkannya. Data ini nantinya dapat digunakan bagi pihak kementerian ESDM dan industri pembangkit tenaga listrik EBT dan non EBT.

Strategi yang dapat diterapkan adalah dengan menggunakan prediktor emisi karbon, pemerintah maupun industri energi listrik dapat lebih sadar dan mempertimbangkan kerusakan lingkungan negara ini akibat produksi listrik,

terkhususnya dari sumber non EBT. Lalu dengan prediktor produksi listrik per sumber, mereka dapat menentukan seberapa kecil emisi karbon yang diinginkan di masa mendatang, data ini bisa didukung dengan data lain seperti populasi yang didapat dari sensus penduduk. Hasil prediksi ini bisa dijadikan *goal* bagi pemerintah dan atau industri pembangkit listrik untuk bisa mengimbangi penggunaan sumber daya EBT dan non EBT dalam rangka menurunkan emisi karbon.

Program EE_predict ini melakukan percobaan dengan model regresi linier dan model kNN. Keduanya memiliki akurasi yang sangat baik, tapi dapat dibuktikan bahwa model regresi linier lebih baik daripada model kNN untuk kasus prediksi *dataset* ini, dengan mengetahui nilai koefisien determinasi dari kedua model tersebut yang mana model regresi linear mempunyai nilai yang lebih besar daripada model kNN.

Jika dilihat dari error relatifnya, prediksi emisi karbon dari listrik dan produksi listrik sumber non EBT sudah cukup akurat, tetapi validasi data ini bergantung pada toleransi error masing-masing instansi yang ingin menggunakannya. Sedangkan untuk prediksi produksi listrik sumber EBT masih memiliki ekor yang besar. Sehingga dapat dikatakan bahwa program ini memiliki suatu kelemahan yang diakibatkan oleh penggunaan *dataset* yang berukuran kecil, kurang lengkap, dan berasal dari sumber yang berbeda-beda.

Untuk pengembangan program kedepannya, dapat dilakukan penelitian lebih lanjut mengenai faktor-faktor yang mempengaruhi produksi listrik dan emisi karbonnya sehingga data yang dipakai lebih akurat dan memiliki keterkaitan yang lebih bagus, serta dengan mencari data yang jauh lebih lengkap.

V. REFERENSI (5 points)

Earthhero. (2019, March 20). What are Carbon Emissions (and why do they matter?). Retrieved April 20, 2020, from <https://earthhero.com/carbon-emissions/>

Mitchell, Tom, "Machine Learning". WCB. 1997

Union of Concerned Scientists. (2017, December 20). Benefits of Renewable Energy Use. Retrieved April 20, 2020, from <https://www.ucsusa.org/resources/benefits-renewable-energy-use>

Worldometers. (n.d.). CO2 Emissions. Retrieved from <https://www.worldometers.info/co2-emissions/>

LAMPIRAN (5 points)

Link Google Drive:

https://drive.google.com/drive/folders/1_ThoNRrnlNWRdXCcejEljnZ_wPLaZYaR?usp=sharing

Link GitHub: https://github.com/hndsmkkgr/EE_Predictor