**Latar Belakang**

Terjadinya bencana alam dapat mempengaruhi baik ekspor maupun impor. Secara umum, ekonomi suatu negara akan terkena dampak dari terjadinya suatu bencana alam. Selain hal tersebut terdapat juga kemungkinan terjadinya efek domino terhadap negara ekspor dan impor di sekitar atau yang berhubungan dengan negara yang terdampak. Hal ini dikarenakan adanya hubungan global antar negara pada era globalisasi ini. Sebagai dampak dari bencana alam adalah kemungkinan naiknya impor dan menurunnya ekspor.

Meningkatnya nilai impor dapat dijelaskan dikarenakan kebutuhan untuk membangun kembali area terdampak. Hal ini dikarenakan bencana alam seperti gampa bumi, tsunami, puting beliung dan sebagainya menyebabkan kerusakan pada infrastruktur publik, menurunnya produksi, serta mengganggu rantai produksi dan distribusi material.

Sebaliknya, menurunnya nilai ekspor dikarenakan menurunnya sumber daya dan energi, kerusakan infrastruktur. Apalagi jika area yang terdampak tersebut merupakan basis produksi barang ekspor, maka akan menurunkan ekspor.

Berdasarkan data dari Centre for Research on the Epidemilogy of Disaster (CRED) selama tahun 2001 sampai dengan 2010, 5 negara yang paling banyak terkena bencana alam adalah China, Amerika Serikat, Filipina, India, dan Indonesia. Indonesia ada di dalam daftar tersebut. Oleh karena itu penting untuk memperhatikan masalah ini dalam menentukan kebijakan masalah ekspor dan impor.

Saat ini telah banyak lembaga yang mengumpulkan dan mengelola data bencana ini, seperti CRED, NGDC, termasuk BNPB di Indonesia. Akan tetapi data tersebut dapat dikatakan tidak real time. Di era digitalisasi sekarang ini terdapat sumber data real time yang dapat diperoleh dalam istilahnya Big Data. Big data sangat berguna dalam memberikan informs yang lebih cepat (real time), akurat dan detail daripada sumber konvensional. Sehingga dapat dipertimbangkan sumber resource baru untuk menganalisa pengaruh bencana alam terhadap ekspor dan impor.

Dikarenakan pertimbangan tersebut, maka penelitian ini dilaksanakan untuk mempelajari penggunaan big data untuk melihat pengaruh bencana alam terhadap ekspor dan impor. Dalam hal ini dari sudut pandang jumlah pemberitaan terkait bencana yang ditangkap oleh GDELT. Global Data on Event Location and Tone database yang menyimpan lebih dari seperempat miliar data mengenai even global mulai dari tahun 1979.

Selain pemberitaan mengenai bencana alam, akan dipelajari juga kejadian-kejadian yang berpengaruh terhadap nilai ekspor dan impor untuk memberikan gambaran yang lebih luas sebagai bahan pertimbangan penentuan kebijakan terkait ekspor impor. Diharapkan nantinya melalui data pemberitaan yang diperoleh dari GDELT, dapat memperkirakan kemungkinan naik maupun turunnya ekspor dan impor sehingga dapat dijadikan bahan pertimbangan kebijakan ekspor impor untuk mengantisipasi hal tersebut.

Pemberitaan mengenai bencana alam akan diekstrak dari database GKG GDELT. Sedangkan kejadian/even lainnya akan diambil dari database Event GDELT.

**Research Questions**

1. Apakah informasi berita yang ditangkap GDELT merepresentasikan situasi yang sebenarnya tentang bencana alam yang terjadi di Indonesia.
2. Apakah terdapat hubungan antara pemberitaan bencana alam Indonesia berpengaruh terhadap naik turunnya ekspor dan impor.
3. Even/kejadian seperti apa yang signifikan berpengaruh terhadap ekspor dan impor.
4. Apakah dapat dibuat sebuah prediksi naik turunnya ekspor impor berdasarkan pemberitaan yang ditangkap GDELT.

**Methodology**

Metodologi yang digunakan dalam proyek ini secara umum melalui tiga tahapan utama, yaitu *Data Understanding*, kemudian *Data Collection, Preprocessing* *and* *Storing* dan *Data Analysis*.

*Data Understanding* meliputi eksplorasi terhadap data pokok yang dijadikan sumber penelitian yaitu GDELT event diikuti dengan pencarian dan eksplorasi terhadap data pendukung yang dibutuhkan untuk melengkapi informasi dari GDELT yang dalam lingkup masalah ini adalah indikator ekonomi makro negara Indonesia (ekspor, impor, nilai tukar rupiah terhadap USD dan nilai IHSG).

*Data Collection, Preprocessing and storing*:

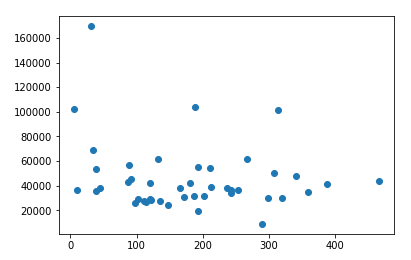
1. Data pemberitaan bencana alam diekstrak dari database GKG menggunakan Tools BigQuery Google.
2. Data bencana di Indonesia diperoleh dari website BNPB.
3. Data ekspor dan impor diperoleh dari website BPS.
4. Data IHSG dan Nilai Tukar rupiah digunakan juga sebagai bahan pendukung berdasarkan teori performa ekonomi yang beberapa diantara indikatornya adalah ekspor-impor, IHSG, dan Nilai Tukar rupiah.
5. Data pemberitaan mengenai even-even yang terjadi di Indonesia diekstrak dari database Events GDELT menggunakan Tools BigQuery Google.

Selanjutnya dilakukan pemilihan fitur-fitur yang relevan di dalam data, *cleansing data* jika diperlukan dan pada akhirnya adalah akan disimpan di dalam satu file atau database (*data integration*) untuk dianalisis lebih lanjut di tahapan berikutnya.

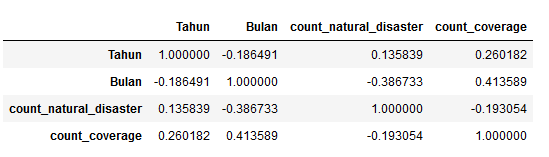
**Question 1**: Apakah informasi berita yang ditangkap GDELT merepresentasikan situasi yang sebenarnya tentang bencana alam yang terjadi di Indonesia.

Hipotesisnya adalah bahwa jumlah pemberitaan mengenai bencana alam akan berkorelasi positif dengan jumlah bencana alam yang sebenarnya terjadi.

Berikut ini gambaran scatter plot hubungan antara jumlah kejadian bencana alam yang sebenarnya terjadi terhadap jumlah pemberitaan yang ada



Dilihat dari scatter plot tersebut, dapat ditarik kesimpulan bahwa korelasi di antara keduanya cukup kecil. Dan berikut adalah perhitungan korelasi pearson:

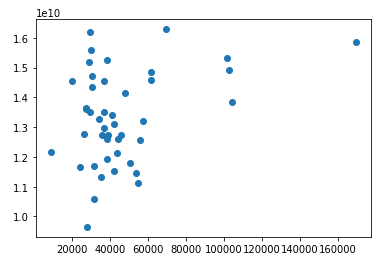


Dapat dilihat bahwa antara jumlah kejadian sebenarnya dengan jumlah pemberitaan justru berbanding terbalik dan dengan angka korelasi yang kecil. Hal ini dapat dijelaskan dengan fakta terjadinya perkembangan teknologi informasi, dimana jumlah pemberitaan yang ada itu semakin banyak dari tahun ke tahun tidak mempertimbangkan banyaknya kejadian.

**Question 2**: Apakah terdapat hubungan antara pemberitaan bencana alam Indonesia berpengaruh terhadap naik turunnya ekspor dan impor.

Hipotesisnya adalah terdapat pengaruh antara jumlah pemberitaan bencana alam Indonesia terhadap ekspor dan impor.

Berikut ini adalah gambaran hubungan antara pemberitaan bencana alam Indonesia terhadap nilai ekspor.



Dilihat dari scatter plot tersebut, dapat ditarik kesimpulan bahwa korelasi di antara keduanya cukup kecil. Dan berikut adalah perhitungan korelasi pearson:



Korelasi positif dengan nilai yang cukup kecil di antara keduanya.

Sehingga tidak cukup menjelaskan ekspor hanya dengan pemberitaan tentang bencana alam Indonesia. Oleh karena itu dilakukan eksplorasi terhadap 20 jenis even berdasarkan event root code CAMEO GDELT untuk menjawab research question ke-3 sebagai berikut.

**Question 3:** Even/kejadian seperti apa yang signifikan berpengaruh terhadap ekspor dan impor

**# Langkah Pertama**

Langkah awal yang kita lakukan ialah menggabungkan keseluruhan data yang telah didapat menjadi data yang saling terintegrasi satu sama lain agar kita dapat melihat perbandingan antara masing-masing data yang terkait. Hasilnya, kita mendapati data yang terdiri dari 237 baris dan 43 kolom yang diantaranya kolom Tahun disatukan dengan Bulan (dengan format YYYYMM), nilai ekspor beserta presentase perubahannya, nilai impor juga beserta presentase perubahannya, CountEvent yang terbagi menjadi 20 jenis, nilai IHSG dengan presentase perubahannya, nilai kurs, dan beberapa kolom lainnya.

**# Langkah Kedua**

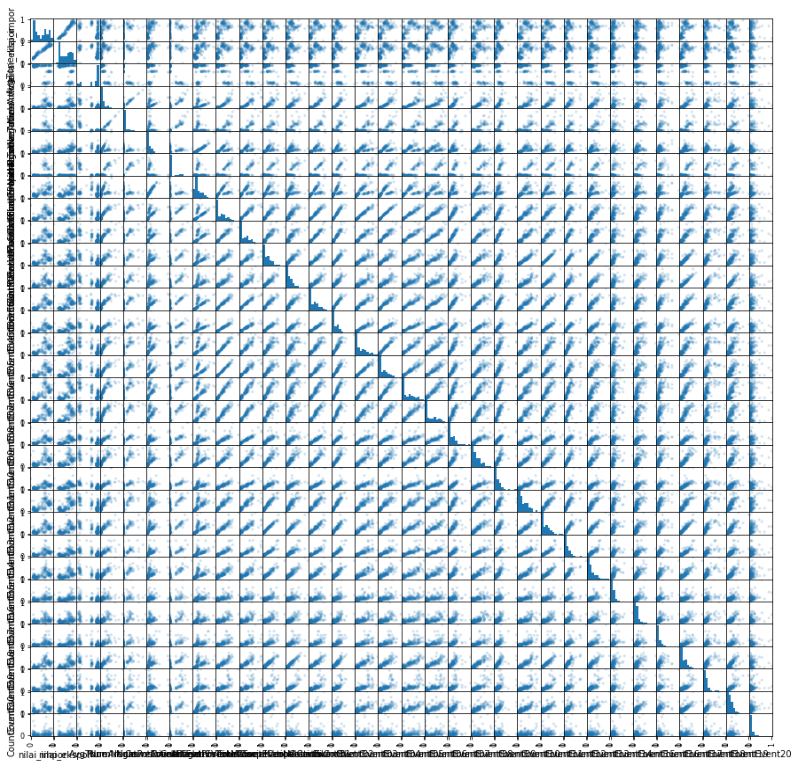
Melakukan *checking* atau koreksi pada tabel data terhadap adanya *missing value* dalam data. Kita ketahui sebelumnya bahwa sebenarnya ada 237 baris untuk masing-masing kolomnya, namun untuk kurs hanya ada 213 baris, sehingga pada kolom Kurs terdapat *missing value* yang kemudian agar tidak mempengaruhi pola maupun tren data yang dihasilkan, maka kita memutuskan untuk mengabaikan baris tersebut sehingga untuk data yang telah kita *cleansing* memiliki 213 baris untuk setiap kolomnya.

**# Langkah Ketiga**

Normalisasi data yang bertujuan untuk menyamakan rentang nilai dari masing-masing variabel data. Kemudian melakukan analisis statistik sederhana dengan menghitung nilai *Mean, Mode, Median, Min, Max, Quartil Q1, Quartil Q3,* serta tidak lupa menghitung standard deviasi dari masing-masing variabel data di setiap kolom. Berikut adalah contoh potongan tabel (hanya menampilkan tiga variabel) yang dilakukan penghitungan statistiknya:

|  |  |  |  |
| --- | --- | --- | --- |
|  | *nilai\_impor* | *nilai\_ekspor* | *AvgTone* |
| *Count* | 239.0000 | 239.0000 | 239.0000 |
| *Mean* | 0.471488 | 0.543056 | 0.739073 |
| *Std* | 0.283955 | 0.237454 | 0.317330 |
| *Min* | 0.000000 | 0.000000 | 0.000000 |
| *25%* | 0.183972 | 0.284851 | 0.661996 |
| *50%* | 0.481752 | 0.552558 | 0.907780 |
| *75%* | 0.722243 | 0.767554 | 0.938910 |
| *Max* | 1.000000 | 1.000000 | 1.000000 |

Selanjutnya lakukan pengecekan terhadap hubungan antar variabel dalam data, berikut adalah gambaran untuk masing-masing korelasi dari setiap data.



**# Langkah Keempat**

Kita hitung korelasi antar variabel untuk melihat variabel mana yang cukup signifikan berpengaruh terhadap perubahan maupun perkembangan sektor nilai ekspor, nilai impor, IHSG, maupun Kurs Rupiah. Namun pada milestone kedua kali ini, secara umum kami tampilkan sekali lagi hanya tabel yang berisikan 3 buah variabel yakni antara nilai impor, nilai impor, dan average tone.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *nilai\_impor* | *nilai\_ekspor* | *AvgTone* |
| *nilai\_impor* | 1.000000 | 0.975979 | -0.431044 |
| *nilai\_ekspor* | 0.975979 | 1.000000 | -0.364119 |
| *AvgTone* | -0.431044 | -0.364119 | 1.000000 |

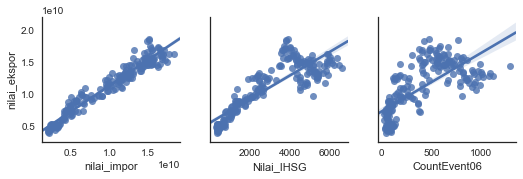
Seperti yang telah disebutkan sebelumnya, maka kita akan melihat variabel yang berpengaruh signifikan terhadap nilai ekspor (yang telah diurutkan dari nilai korelasi yang terbesar).

|  |
| --- |
| nilai\_ekspor                         1.000000 |
| nilai\_impor                          0.972838 |
| Nilai\_IHSG                         0.863856 |
| CountEvent06                         0.703341 |
| CountEventPositiveTone               0.663697 |
| CountEvent05                         0.661689 |
| CountEvent10                         0.653656 |
| CountEventVerbalCooperation          0.644572 |
| CountEventMaterialCooperation        0.641131 |
| CountEvent01                         0.640677 |
| CountEvent04                         0.632503 |
| CountEvent03                         0.631822 |
| CountEvent02                        0.607347 |
| CountEvent07                         0.598671 |
| NumPositiveArticle                   0.596291 |
| CountEvent08                         0.584833 |
| CountEventVerbalConflict             0.577974 |
| CountEvent11                         0.574758 |
| CountEvent17                         0.572598 |
| CountEventMaterialConflict           0.525588 |
| CountEvent09                         0.523753 |
| CountEvent12                         0.490360 |
| NumArticle                           0.475059 |
| CountEvent13                        0.470281 |
| CountEvent14                         0.450466 |
| CountEvent19                         0.437051 |
| CountEvent16                         0.431140 |
| Nilai\_Kurs\_Jual                      0.358183 |
| CountEvent18                         0.354689 |
| CountEvent15                         0.329466 |
| CountEventNegativeTone               0.265008 |
| CountEvent20                         0.253547 |
| NumNegativeArticle                   0.237258 |
| persentase\_perubahan\_nilai\_ekspor    0.102541 |
| persentase\_perubahan\_nilai\_impor     0.085683 |
| Persentase\_Kurs                      0.085440 |
| Persentase\_IHSG                     -0.125854 |
| AvgTone                             -0.383984 |

Dari sorting korelasi tersebut diperoleh 3 variabel yang memiliki nilai korelasi > 0.7 terhadap nilai ekspor, yaitu: jumlah kejadian event 06, nilai IHSG dan nilai Import. sehingga untuk selanjutnya akan kita cari model yang menjelaskan hubungan tersebut dengan nilai ekspor sebagai y (variabel dependen) dan count event 06, nilai IHSG dan nilai Impor sebagai variabel independen x1, x2, dan x3.

**# Langkah Kelima**

Melakukan analisis statistik regresi liniernya.



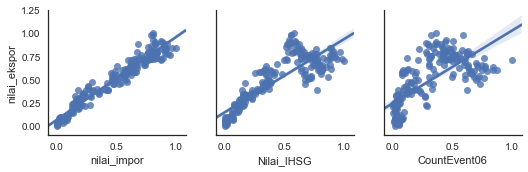
Dengan menggunakan bantuan python maka kita dapatkan model yang terbentuk untuk regresi linier yakni dengan persamaan:

y = 3.23e+09 + (7.61e-01)nilai\_impor + (1.92e+05)Nilai\_IHSG - (5.56e+05)\*CountEvent06.

Sehingga kita ketahui bahwa nilai impor dan nilai IHSG berkorelasi positif terhadap nilai ekspor, jika nilai tersebut naik maka nilai ekspor akan naik namun jika turun maka nilai ekspor akan turun. Sementara itu CountEvent06 berkorelasi negatif terhadap nilai ekspor, jika jumlah event06 naik maka nilai ekspor turun dan sebaliknya.

**# Langkah keenam**

Melakukan evaluasi model yang didapat menggunakan *Mean absolute error (MAE), Mean squared error (MSE), or Root mean squared error (RMSE)*. Namun sebelumnya kita lakukan standardisasi untuk masing-masing variabelnya lalu lakukan kembali analisis statistik regresi linier untuk data yang telah distandardisasi ini, maka akan didapat plot seperti dibawah ini:



model yang terbentuk ialah:

**y = 0.0066 - (0.9113)nilai\_impor - (0.0048)Nilai\_IHSG - (0.0087)\*CountEvent06**

artinya nilai impor berkorelasi positif terhadap nilai ekspor, jika nilai tersebut naik maka ekspor naik namun jika turun maka ekspor turun. Sementara itu nilai IHSG dan CountEvent06 berkorelasi negatif terhadap nilai ekspor, jika jumlah event06 naik maka nilai ekspor turun dan sebaliknya.

Didapatkan pula hasil dari masing-masing nilai evaluasi dan analisis statistik terhadap perubahan nilai ekspor yang telah distandardisasi, yakni:

MAE = 0.0514

MSE = 0.0043

RMSE = 0.0662

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *nilai\_ekspor* | *nilai\_impor* | *Nilai\_IHSG* | *CountEvent06* |
| *count* | 213.000000 | 213.000000 | 213.000000 | 213.000000 |
| *mean* | 0.470958 | 0.452018 | 0.408943 | 0.287625 |
| *Std* | 0.279739 | 0.305550 | 0.307049 | 0.251537 |
| *Min* | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| *25%* | 0.223044 | 0.154337 | 0.115336 | 0.050753 |
| *50%* | 0.516685 | 0.499233 | 0.366431 | 0.222839 |
| *75%* | 0.721374 | 0.737608 | 0.686302 | 0.486915 |
| *Max* | 1.000000 | 1.000000 | 1.000000 | 1.000000 |

Interpretasinya, dengan nilai RMSE 0.06 dibandingkan range data normalize adalah 0-1 maka dapat dikatakan bahwa nilai RMSE nya cukup kecil sehingga model yang cocok dihasilkan:

**y = 0.0066 - (0.9113)nilai\_impor - (0.0048)Nilai\_IHSG - (0.0087)\*CountEvent06**

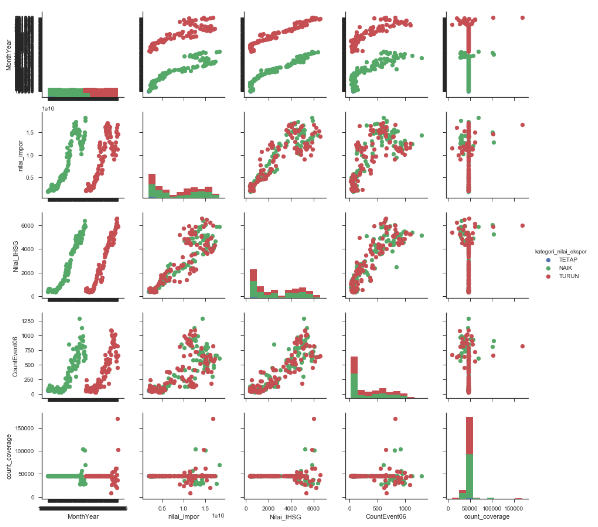
cukup fit menggambarkan hubungan nilai ekspor dengan impor, IHSG, dan count event 06.

**Question 4**: Apakah dapat dibuat sebuah prediksi naik turunnya ekspor impor berdasarkan pemberitaan yang ditangkap GDELT

Response variable: kondisi naik/turun nilai ekspor (data kategorik)

Independent variable: nilai impor, nilai IHSG, jumlah kejadian Event root code 06, jumlah pemberitaan tentang bencana alam (keempat variabel ini bertipe numerik)

Untuk memprediksi naik/turunnya nilai ekspor jika nilai impor, nilai IHSG, jumlah kejadian Event root code 06 dan jumlah pemberitaan tentang bencana alam akan digunakan machine learning dengan metode KNN (K Near Neighbors). Hal ini dikarenakan kategori yang akan diprediksi sudah diketahui yaitu naik (1) atau turun (0) nilai ekspor. Dan sudah terdapat data dengan label naik maupun turun. Dalam hal ini dilakukan penambahan feature pada dataset nilai ekspor yaitu kategori\_nilai\_ekspor yang diisikan berdasarkan naik atau turunnya nilai ekspor pada bulan n terhadap nilai ekspor pada bulan n-1. Metode tersebut dipilih juga dikarenakan bentuk sebaran sebagai berikut:

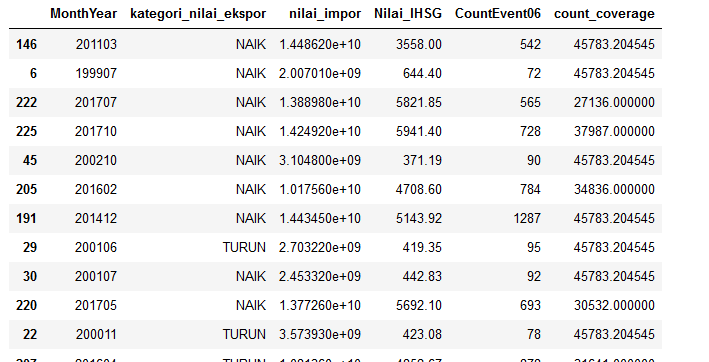


KNeighborsClassifier(algorithm='auto', leaf\_size=30, metric='minkowski',

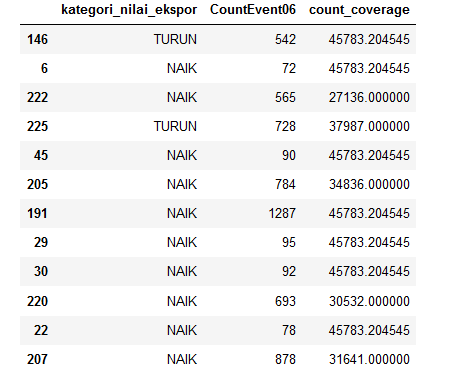
metric\_params=None, n\_jobs=1, n\_neighbors=10, p=2,

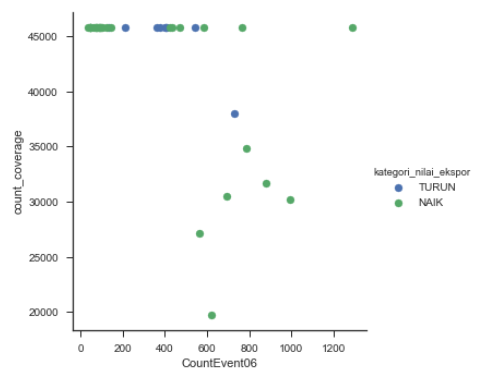
weights='distance')

Ambil 40 sampel:



Prediksi kategori\_nilai\_ekspor dari sampel:





Hitung akurasi classification:

