**PRAPROSES DATA**

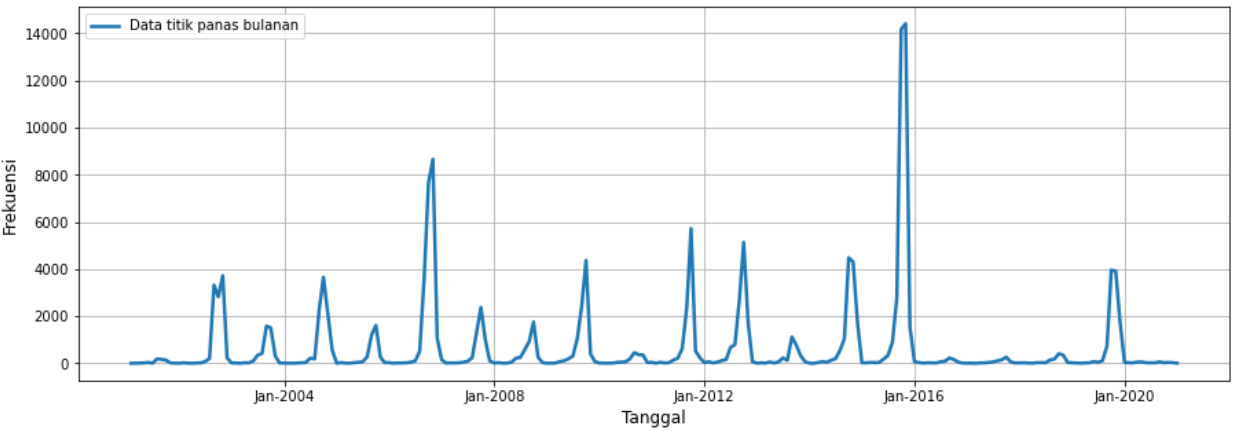
Setelah berhasil mendapatkan data penelitian dari FIRMS NASA dan NOAA, selanjutnya dilakukan praproses data. Praproses data merupakan suatu proses untuk mempersiapkan data-data yang telah dikumpulkan menjadi lebih terstruktur. Pada penelitian ini, praproses data yang dilakukan adalah seleksi atribut, aggregasi data, integrasi data, dan normalisasi data.

1. Seleksi atribut

Data titik panas dari FIRMS NASA memiliki 15 atribut. Hasil seleksi atribut hanya menggunakan 4 atribut yaitu *latitude, longtitude, acq\_date, confidence*. Karena keempat atribut ini sudah mewakili kemunculan titik panas.

1. Aggregasi data

Jumlah titik panas di Provinsi Sumatra Selatan dari bulan Januari 2001 sampai Desember 2020 sebanyak 155.661 titik panas. Hal ini disebabkan data titik panas dari FIRMS NASA adalah data harian. Sedangkan data SST *Nina* 3.4, SOI, ONI dari bulan Januari 2001 sampai Desember 2020 sebanyak 240 data. Karena data SST *Nina* 3.4, SOI, ONI berbentuk data bulanan. Oleh karena itu, data titik panas dilakukan aggregasi data untuk merubah data harian menjadi data bulanan. Hasil aggregasi data titik panas ditunjukan pada Gambar 2.



Gambar 2. Data titik panas bulanan Provinsi Sumatra Selatan

1. Integrasi data

Data titik panas, SST Nina 3.4, SOI, ONI merupakan data yang saling berdiri sendiri. Namun pada penelitian ini data titik panas menjadi variabel terikat dan data SST Nina 3.4, SOI, ONI menjadi variabel bebas. Oleh karena itu, dilakukan integrasi data untuk menggabungkan data-data tersebut. Hasil integrasi data seperti ditunjukan pada Tabel 2.

Tabel 2 Contoh hasil integrasi data titik panas, SST Nina 3.4, SOI, ONI

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tahun | Bulan | SST Nina 3.4 | SOI | ONI | Titik panas |
| 2020 | 1 | 0,53 | 0,3 | 0,50 | 27 |
| 2020 | 2 | 0,42 | -0,1 | 0,48 | 24 |
| 2020 | 3 | 0,61 | -0,2 | 0,40 | 59 |
| 2020 | 4 | 0,54 | 0,3 | 0,19 | 60 |
| 2020 | 5 | -0.26 | 0,7 | -0.08 | 27 |

1. Normalisasi data

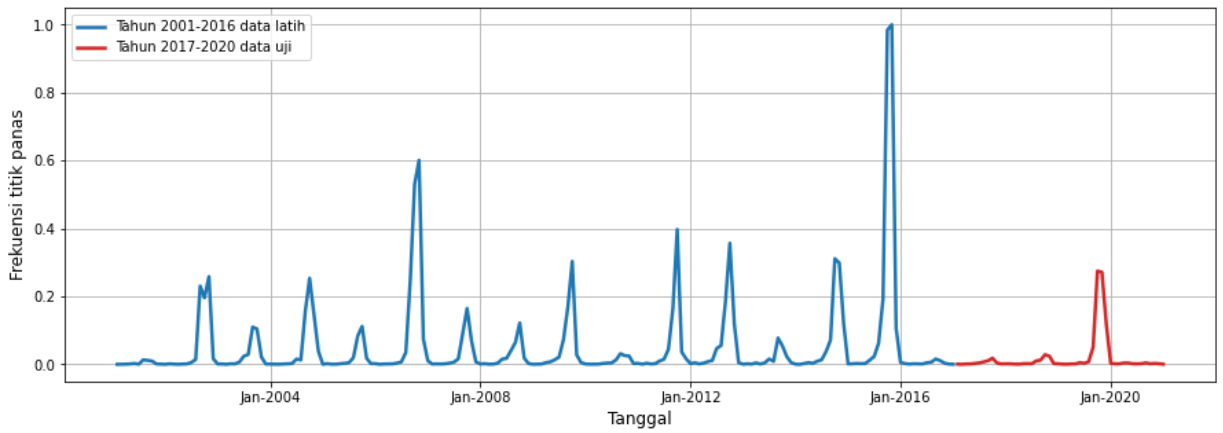
Data titik panas, SST Nina 3.4, SOI, ONI memiliki satuan yang berbeda-beda. Oleh karena itu dilakukan normalisasi data untuk menyamakan satuan datanya. Normalisasi data menggunakan metode max-min. Hasil normalisasi data seperti ditunjukan pada Tabel 3.

Tabel 3 Contoh hasil normalisasi data titik panas, SST Nina 3.4, SOI, ONI

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Tahun | Bulan | SST Nina 3.4 | SOI | ONI | Titik panas |
| 2020 | 1 | 0,5000 | 0,5500 | 0,5000 | 0,0018 |
| 2020 | 2 | 0,4773 | 0,5100 | 0,4953 | 0,0016 |
| 2020 | 3 | 0,5165 | 0,5000 | 0,4766 | 0,0040 |
| 2020 | 4 | 0,5021 | 0,5500 | 0,4276 | 0,0041 |
| 2020 | 5 | 0,3368 | 0,5900 | 0,3645 | 0,0018 |

**Pembagian Data**

Pembagian data merupakan proses untuk membagi data menjadi data latih dan data uji. Data latih digunakan untuk proses pembelajaran mesin model prediksi dan data uji digunakan untuk melakukan validasi dari hasil pembelajaran mesin. Pembagian data menggunakan metode *split validation* dengan perbandingan 80% data latih dan 20% data uji. Sehingga periode deret waktu data latih dari tahun 2001 sampai 2016 dan data uji dari tahun 2017 sampai 2020. Hasil pembagian data seperti ditunjukan pada Gambar 3.



Gambar 3 Hasil pembagian data latih dan data uji