

## Manajemen Komunikasi Strategis

# TUTORIAL PENGGUNAAN PLATFORM ORANGE UNTUK DATA MINING DAN ANALISA SENTIMEN

**Laboratorium Public Relations** 

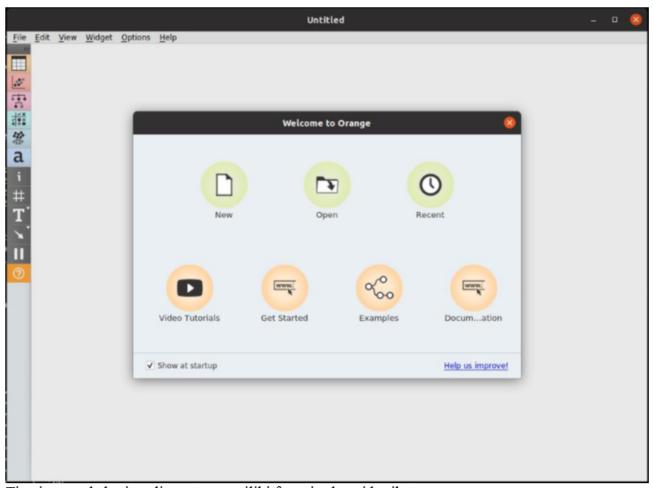


Disusun oleh Dr.Indar Sugiarto

## Penggunaan Orange Dasar

#### 1.1. Tampilan Awal

Pada saat pertama kali Orange dijalankan, akan muncul tampilan awal sebagai berikut:



Tiga icon pada baris paling atas memiliki fungsi sebagai berikut:

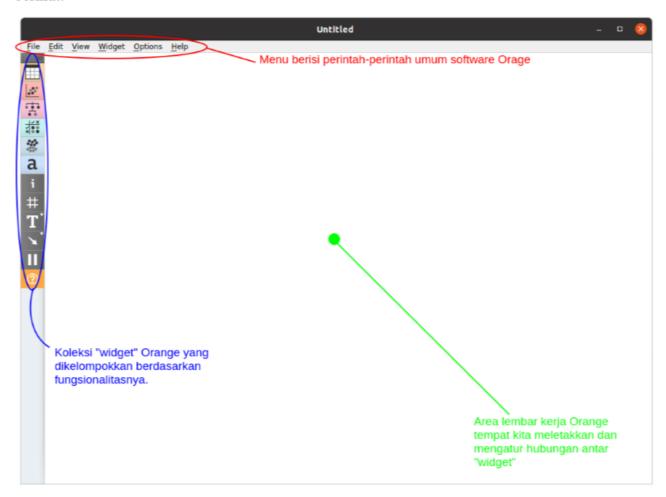
- New → membuat lembar kerja baru
- Open → membuka lembar kerja Orange yang pernah dibuat sebelumnya
- Recent → membuka lembar kerja Orange yang terakhir kali dibuka (akan dimunculkan daftarnya)

Sedangkan baris-baris paling bawah berisi informasi-informasi terkait penggunaan software Orange tersebut.

Untuk membuat lembar kerja Orange yang baru, silahkan klik pada icon "New".

#### 1.2. Bekerja dalam Lembar Kerja Orange

Berikut adalah tampilan lembar kerja (disebut "canvas") yang masih kosong ketika Orange dibuka dengan perintah "New". Dalam lembar kerja ini nanti kita akan membuat "workflow", yaitu model aliran dan pemrosesan data dalam Orange. Bagian-bagian lembar kerja Orange adalah sebagai berikut:

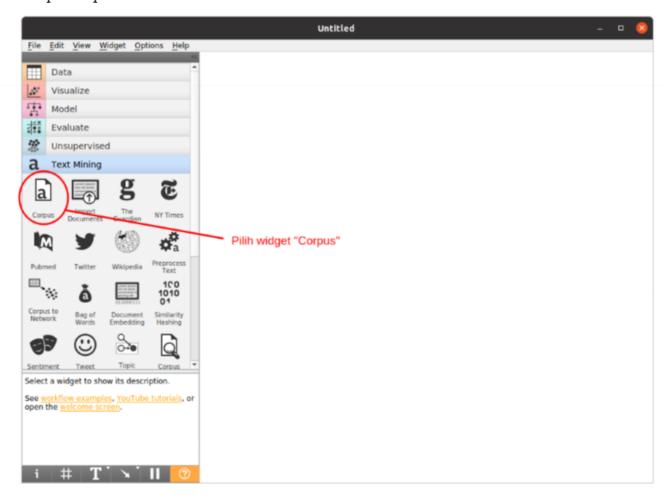


Perhatikan di bagian kiri terdapat "Tool Dock" yang berisi "widget-widget" Orange. Widget adalah komponen Orange untuk menjalankan fungsi/komputasi tertentu, misalkan: membaca file data, menampilkan data dalam bentuk tabel, melakukan regresi, dan sebagainya.

Orange dilengkapi dengan banyak sekali contoh-contoh yang menjelaskan beragam alur kerja machine learning (data mining). Dalam modul-1 ini kita akan belajar hal dasar yaitu menampilkan data hasil text mining dalam bentuk word cloud.

#### 1.3. Bekerja dengan Workflow di Orange

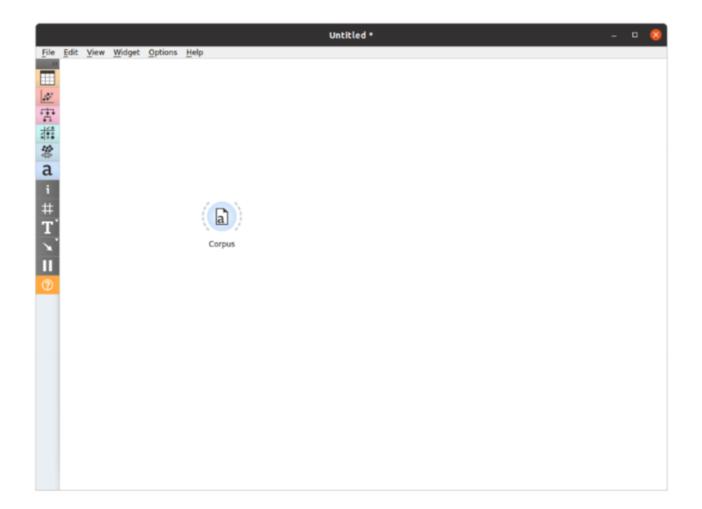
Pada bagian "Tool Dock", aktifkan pada kelompok "Text Mining", lalu klik pada icon berlabel "Corpus" seperti contoh berikut:



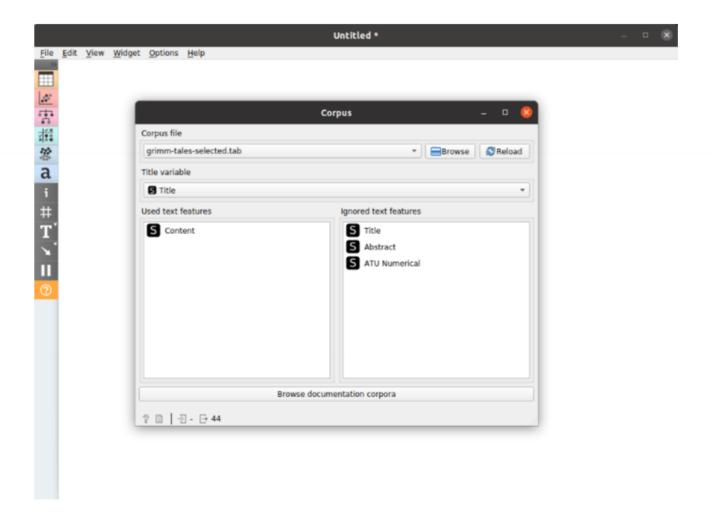
Bisa juga dilakukan dengan cara klik tombol mouse sebelah kanan pada area canvas yang kosong hingga muncul "pop-up box" dimana kita bisa mencari widget yang akan kita pakai:



Setelah itu widget Corpus akan diletakkan di canvas kita. Tampilannya seharusnya seperti ini:

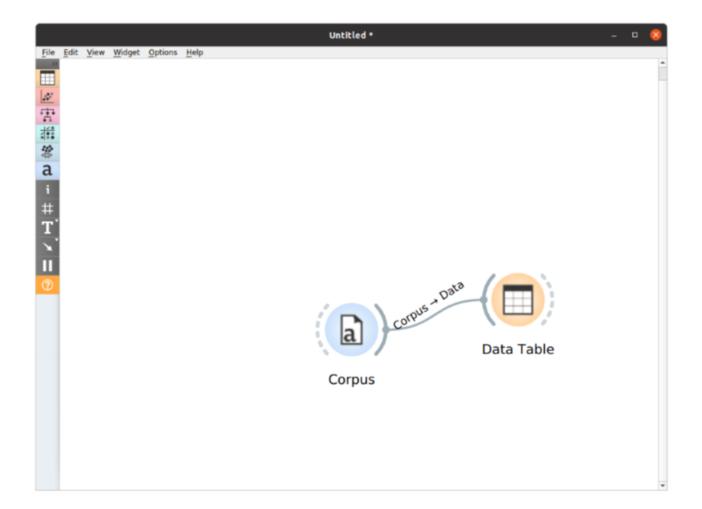


Silahkan double-click pada icon Corpus tersebut hingga muncul Dialog yang berisi daftar corpus yang tersedia di dalam database software Orange. Corpus dalam konteks text mining adalah kumpulan dokumen/naskah berbentuk teks yang masih mentah tetapi sudah disimpan dalam format tertentu (dalam Orange, salah satu format data yang bisa digunakan adalah .tab). Berikut adalah tampilan Dialog Corpus:

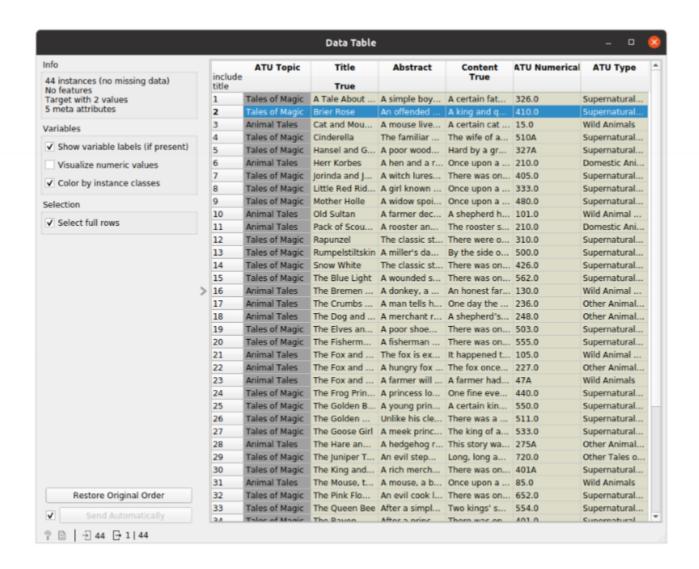


Untuk contoh ini, silahkan pilih "grimm-tales-selected.tab" pada bagian "Corpus file" dalam Dialog Corpus tersebut. Jika pilihan ini diganti-ganti, perhatikan perubahan yang terjadi pada bagian "Title variable", "Used text features", maupun "Ignored text features". Bagian-bagian ini adalah hasil ekstrasi dan interpretasi software Orange terhadap dataset yang dipilih (jadi isinya bisa berbedabeda tergantung dataset yang dipilih). Untuk mengakhiri pemilihan corpus ini, cukup dengan menekan tombol "X" pada pojok kanan atas.

Dalam contoh ini, kita juga akan menampilkan isi dari corpus "grimm-tales-selected". Caranya adalah dengan memilih widget "Data Table" dari "Dock Tools → Data" atau dengan klik tombol kanan mouse lalu tulis/pilih widget "Data Table". Setelah widget Data Table diletakkan di canvas kita, silahkan tarik garis penghubung dari widget "Corpus" ke widget "Data Table" seperti contoh berikut:

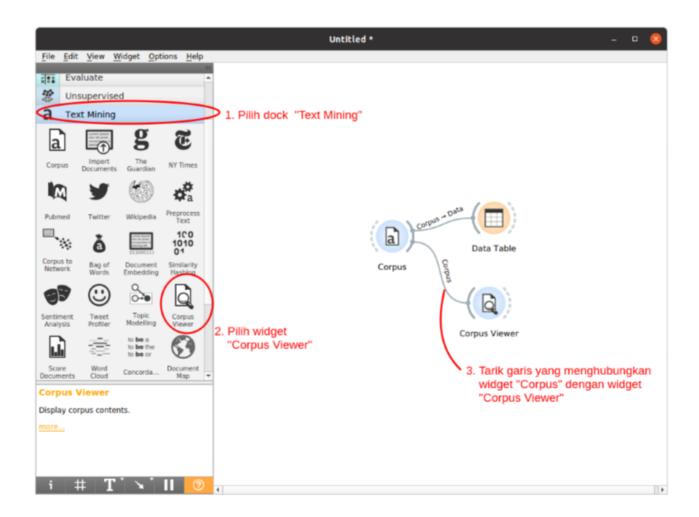


Jika kita double-click widget "Data Table", maka akan ditampilkan isi dari corpus "grimm-tales-selected" dalam bentuk tabel sebagai berikut:

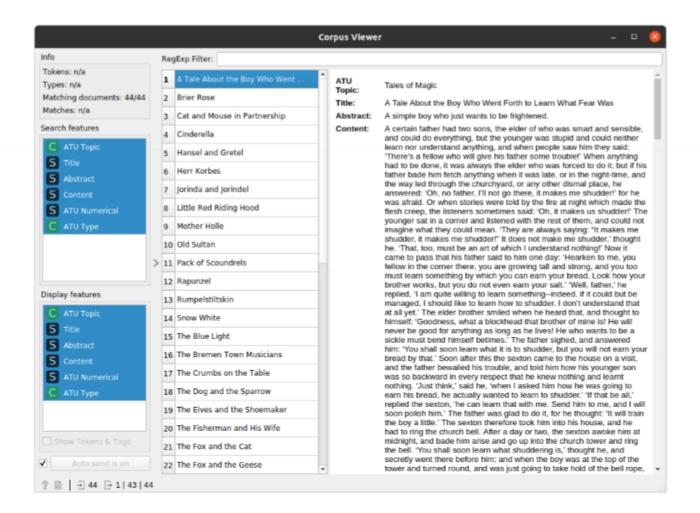


Untuk menutup tampilan tabel di atas, silahkan klik "X" pada pojok kanan atas.

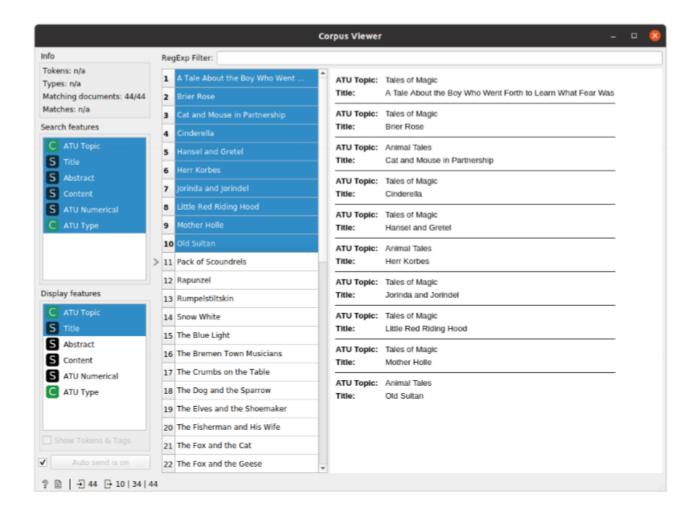
Selain menampilkan isi corpus dalam bentuk tabel seperti di atas, Orange juga menyediakan tampilan khusus untuk corpus yang bisa diakses dengan menggunakan widget "Corpus Viewer". Widget ini bisa diambil dari "Dock Tools → Text Mining" atau klik kanan tombol mouse lalu ketik/pilih "Corpus Viewer". Setelah widget "Corpus Viewer" ini diletakkan di canvas, silahkan tarik garis yang menghubungkan widget "Corpus" dengan widget "Corpus Viewer" seperti contoh berikut:



Silahkan double-click pada widget "Corpus Viewer" untuk menampilkan isi corpus-nya seperti ditunjukkan pada gambar berikut:

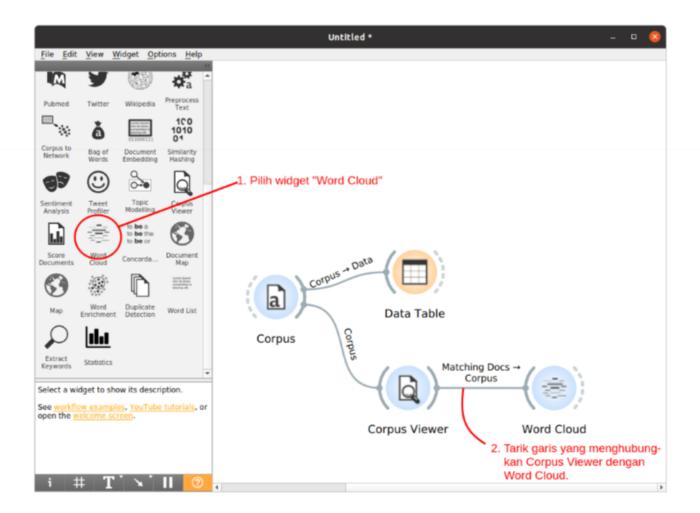


Tampilan dari corpus bisa diatur sedemikian rupa sehingga hanya data-data tertentu saja yang ditampilkan. Contohnya, pada bagian "Display feature" pilih hanya item "ATU Topic" dan "Title" saja (yaitu dengan menekan tombol shift keybord sambil menekan tombol kiri mouse), dan pada daftar dokumen, silahkan pilih dokumen nomor 1 hingga 10 (juga dengan menekan tombol shift keyboard sambil menekan tombol kiri mouse). Maka akan didapatkan tampilan sebagai berikut:



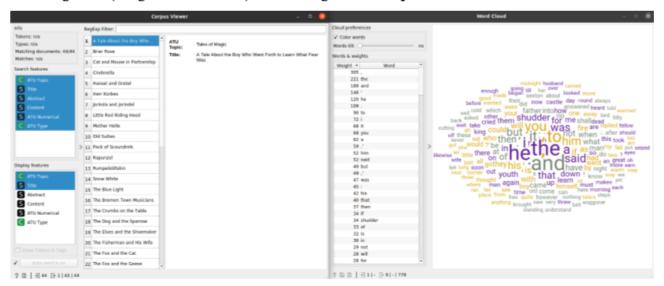
Perhatikan bagaimana Orange menampilkan sebagian data saja dalam window Corpus Viewer di atas. Untuk menutup tampilan Corpus Viewer tersebut, silahkan klik "X" pada pojok kanan atas.

Untuk menampilkan word cloud dari corpus yang kita pilih di atas, silahkan pilih widget "Word Cloud" dari "Dock Tools → Word Cloud" (atau gunakan tombol kanan mouse lalu cari/pilih widget Word Cloud). Setelah widget "Word Cloud" diletakkan pada canvas, silahkan hubungan widget "Word Cloud" tersebut dengan widget "Corpus Viewer" seperti contoh berikut:



Catatan: widget "Word Cloud" bisa juga dihubungkan dengan widget "Corpus". Tetapi dalam percobaan ini, widget "Word Cloud" kita hubungkan dengan widget "Corpus Viewer" supaya hasil tampilan word cloud-nya nanti bisa kita ubah-ubah dengan cara memilihnya dari dalam widget "Corpus Viewer".

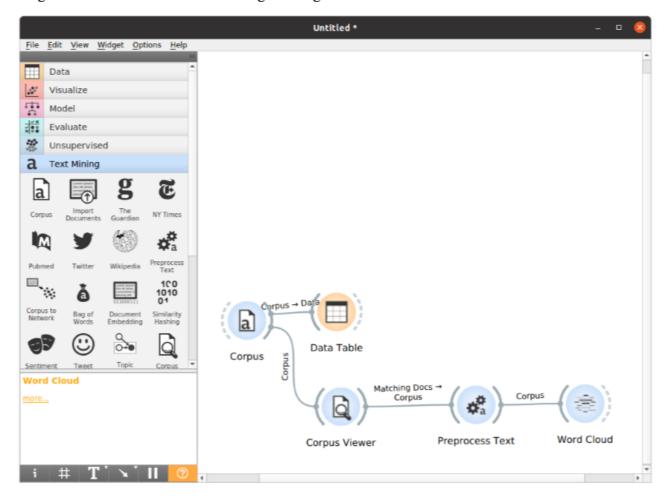
Sekarang buka (dengan double-click) kedua widget ini: "Corpus Viewer" dan "Word Cloud".



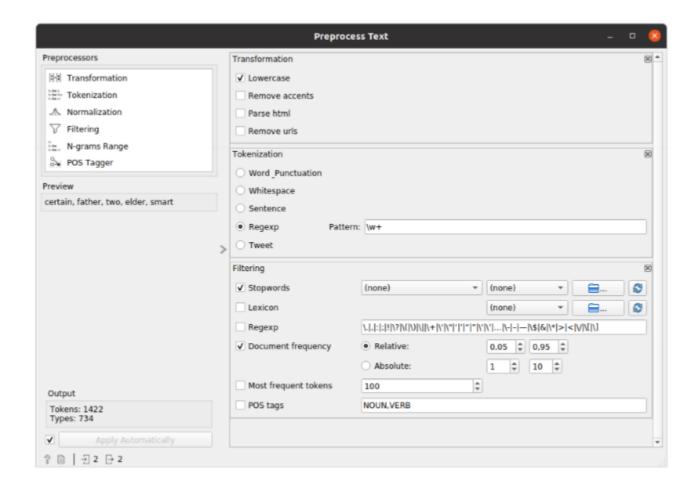
Silahkan ganti-ganti pilihan dokumen pada widget "Corpus Viewer" lalu amati apa yang terjadi dengan word cloud yang dihasilkan!

Sampai di sini, hasil yang didapatkan oleh word cloud kurang tepat karena dia juga menampilkan kata-kata yang sebenarnya tidak relevan seperti kata "the" dan juga simbol-simbol tanda baca. Kata-kata seperti "the", "a", "an", dan sebagainya ini disebut "stop word" dalam dunia NLP dan bisa dihilangkan dari word cloud dengan cara sebagai berikut.

Dari "Dock Tools → Text Mining", pilihlah widget "Preprocess Text" (atau cari/pilih menggunakan pop-up menu dari tombol kanan mouse). Letakkan di canvas di antara widget "Corpus Viewer" dan widget "Word Cloud" lalu buat hubungan sebagai berikut:



Double-click widget "Preprocess Text" untuk mengaktifkan konfigurasinya lalu pilih opsi-opsi seperti ditunjukkan pada gambar berikut:



Pada percobaan kali ini, cukup pastikan pilih/aktifkan opsi sebagai berikut:

- Transformation → Lowercase
- Tokenization → Regexp dengan Pattern "\w+"
- Filtering → Stopwords
- Filtering → Doument frequency → Relative, antara 0.1 hingga 0.9

Untuk menutup jendela konfigurasi tersebut, silahkan klik "X" pada ujung kanan atas. Setelah itu buka lagi tampilan dari widget "Corpus Viewer" dan widget "Word Cloud". Sambil memilih dokumen-dokumen dari Corpus Viewer, perhatikan word cloud yang dihasilkan (seharusnya lebih bersih dari kata-kata yang tidak perlu seperti "the", "an", dan sebagainya).



## **Sentiment Analysis**

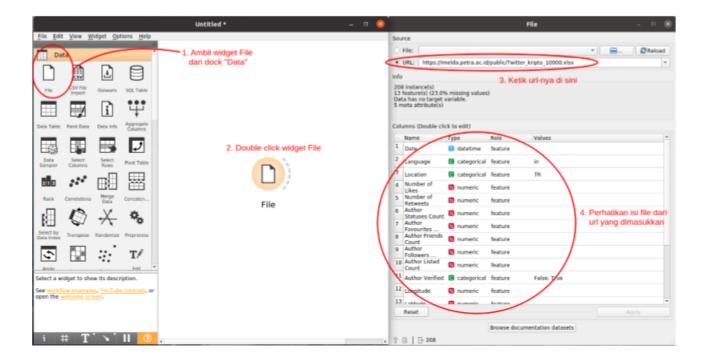
#### 1.1. Sumber Data

Dalam percobaan ini, kita akan melakukan *sentiment analysis* terhadap data-data dari Twitter dengan Twit #**kripto** 

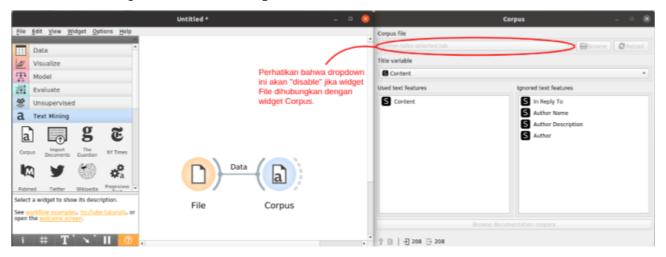
Sebenarnya Orange menyediakan widget untuk bisa mengambil data langsung dari Twitter. Tapi untuk bisa menggunakan widget ini, pengguna harus memiliki Twitter API Key dan Secret-nya yang bisa diperoleh dengan mendaftar terlebih sebagai Twitter developer (lihat https://developer.twitter.com/en/docs/twitter-api/getting-started/getting-access-to-the-twitter-api). Karena tidak semua orang bisa mendapatkannya, maka dalam percobaan ini, data Twitter akan disediakan untuk diakses oleh praktikan.

File yang disediakan adalah file Twitter\_kripto\_10000.xlsx yang sebelumnya sudah diambil dari Twitter per tanggal 10 November 2021. File ini dapat diambil dari url berikut: https://imelda.petra.ac.id/public/Twitter\_kripto\_10000.xlsx

Silahkan jalankan Orange dan buka Lembar Kerja kosong. Tambahkan widget File dari dock "Data". Setelah itu *double click* widget File tersebut dan isikan link dari file Twitter\_kripto\_10000.xlsx sebagai berikut.



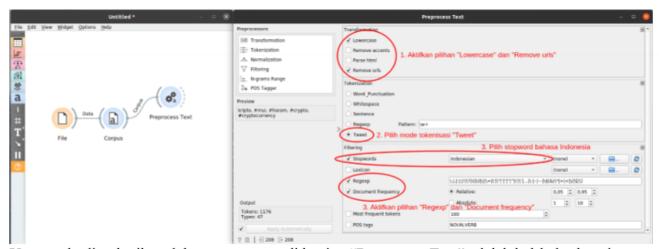
Selanjutnya, tambahkan widget Corpus dari dock "Text Mining" lalu hubungkan widget File di atas ke widget Corpus. Setelah terhubung, silahkan *double-click* widget Corpus dan pastikan bahwa dropdown di dalam Corpus menjadi "disable" (tidak bisa di-klik atau diganti). Ini menandakan bahwa sumber corpus diambil dari widget File.



Seperti sebelumnya, kita bisa menampilkan word cloud dari corpus kita untuk melihat kata apa yang paling dominan dalam corpus tersebut.

Yang perlu diperhatikan adalah, jika kita langsung menampilkannya menggunakan widget "Word Cloud", maka tampilannya akan cukup kacau karena data-data dari Twitter seringkali menggunakan simbol-simbol (seperti tanda baca) yang tidak relevan. Untuk itu kita bisa memprosesnya terlebih dahulu guna menghilangkan simbol atau kata yang tidak relevan tersebut.

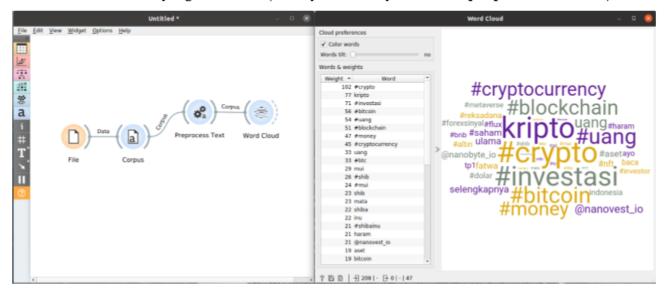
Tambahkan widget "Preprocess Text" dari dock "Text Mining" dan hubungkan dengan output dari widget Corpus. Kemudian *double-click* pada widget "Proprocess Text" untuk mengaturnya sebagai berikut.



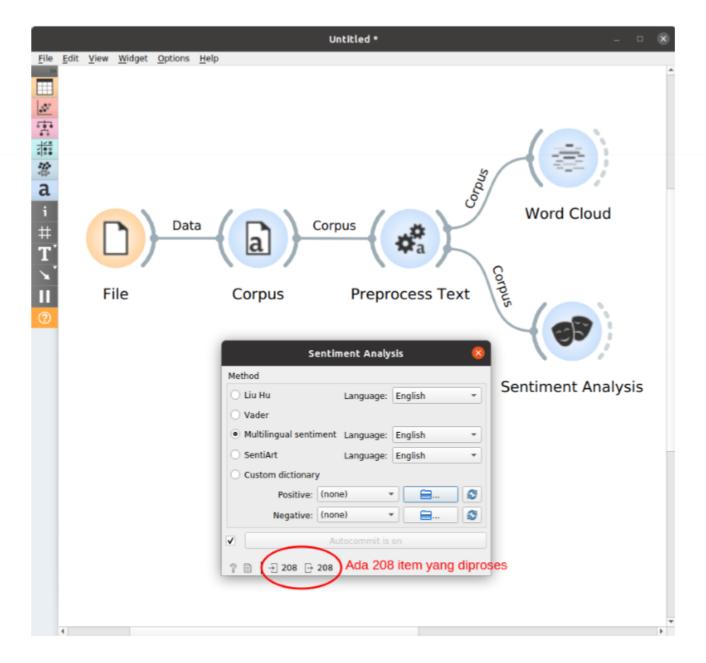
Yang perlu diperhatikan dalam pengaturan di bagian "Preprocess Text" adalah hal-hal sebagai berikut:

- Aktifkan pilihan "Lowercase" dan "Remove urls" pada bagian Transformation.
- Aktifkan pilihan "Tweet" pada bagian Tokenization
- Aktifkan pilihan Stopwords Indonesian pada bagian Filtering
- Aktifkan pilihan "Regexp" dan "Document frequency" pada bagian Filtering

Perhatikan word cloud yang dihasilkan (hasilnya seharusnya sama/mirip seperti di bawah ini).



Selanjutnya kita akan mulai proses sentiment analysis pada corpus yang kita gunakan. Tambahkan widget "Sentiment Analysis" dari dock "Text Mining" dan hubungkan dengan output dari widget Corpus. Lanjutkan dengan *double-click* widget "Sentiment Analysis" tersebut lalu pilihkan metode "Multilingual sentiment" dan pilihlah bahasa Indonesia dari menu dropdown di sebelahnya.

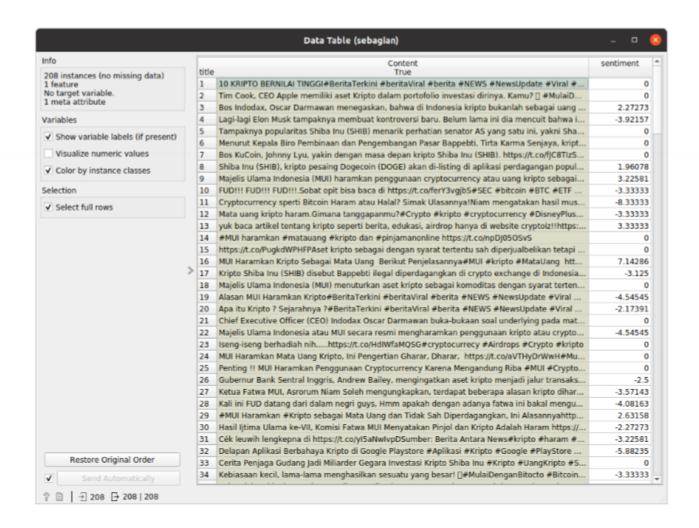


Perhatikan ada 208 item yang diproses oleh widget "Sentiment Analysis" (perhatikan angka yang ditampilkan di bagian bawah dari menu Sentiment Analysis). Untuk melihat kesemua 208 item tersebut secara lengkap, silahkan tambahkan widget "Data Table" dan hubungkan dengan widget "Sentiment Analysis" tersebut. Silahkan *double-click* widget "Data Table" untuk melihat isinya. Isi yang ditampilkan dalam widget "Data Table" terlalu banyak dan tidak semua berguna untuk analisa.

Untuk memilih bagian mana saja yang akan ditampilkan, gunakan widget "Select Columns" yang bisa diambil dari dock "Data". Hubungkan widget "Select Columns" tersebut dengan widget "Sentiment Analysis", lalu *double-click* widget tersebut untuk mengatur fitur mana yang akan di analisa lebih detil.

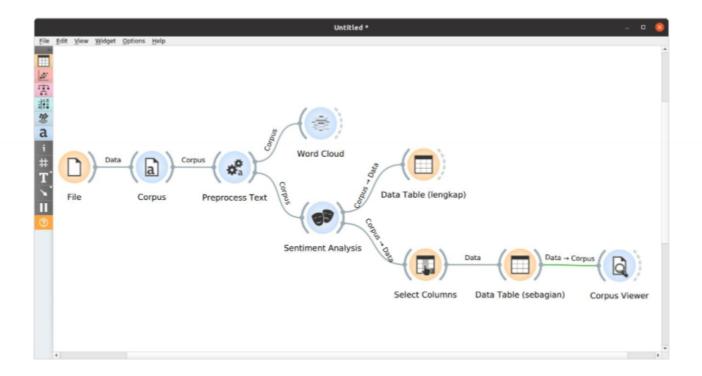


Tambahkan satu lagi widget "Data Table" untuk dihubungkan dengan widget "Select Colums". Jika widget "Data Table" yang baru ini diaktifkan (dengan *double-click*), maka akan ditampilkan hanya "content" dan nilai sentimen-nya saja.

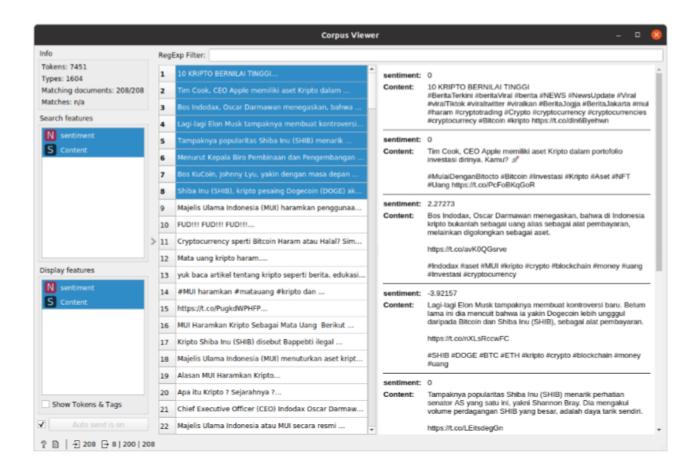


Perhatikan kolom kedua pada tampilan "Data Table" di atas. Kolom ini berisi informasi hasil analisa sentimen dari widget "Sentiment Analysis" dengan mode Multilingual berbahasa Indonesia. Nilai sentimen yang dihasilkan bisa positif (dengan derajat tertentu), bisa juga negatif (juga dengan derajat tertentu), atau nol yang berarti netral.

Untuk melihat nilai sentimen masing-masing cuitan ("Tweet"), tambahkan widget "Corpus Viewer" dan hubungkan dengan "Data Table" sebelumnya.

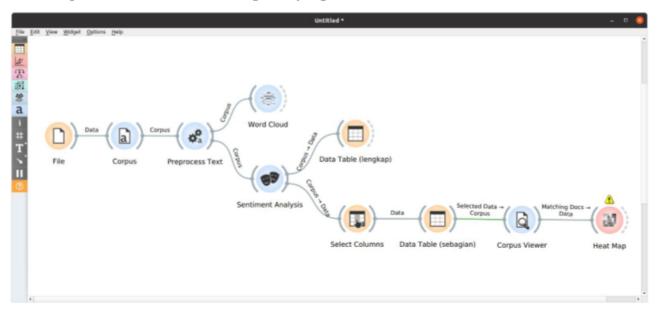


Silahkan double-click pada widget "Corpus Viewer" untuk melihat nilai sentimen masing-masing cuitan. Saat jendela "Corpus Viewer" ditampilkan, kita bisa memilih satu-persatu cuitannya, atau memilih beberapa cuitan sekaligus untuk melihat nilai sentimen-nya.

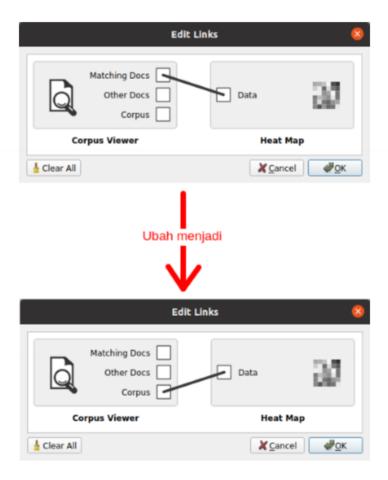


Pada contoh di atas, perhatikan cuitan nomor 3 memiliki sentimen positif, sedangkan cuitan nomor 4 memiliki cuitan negatif. Bisa dilihat bahwa cuitan nomor 1 dan 2 tidak memiliki sentimen karena pada dasarnya kedua cuitan tersebut hanya berisi informasi belaka dan tidak ada pesan subyektif di dalamnya.

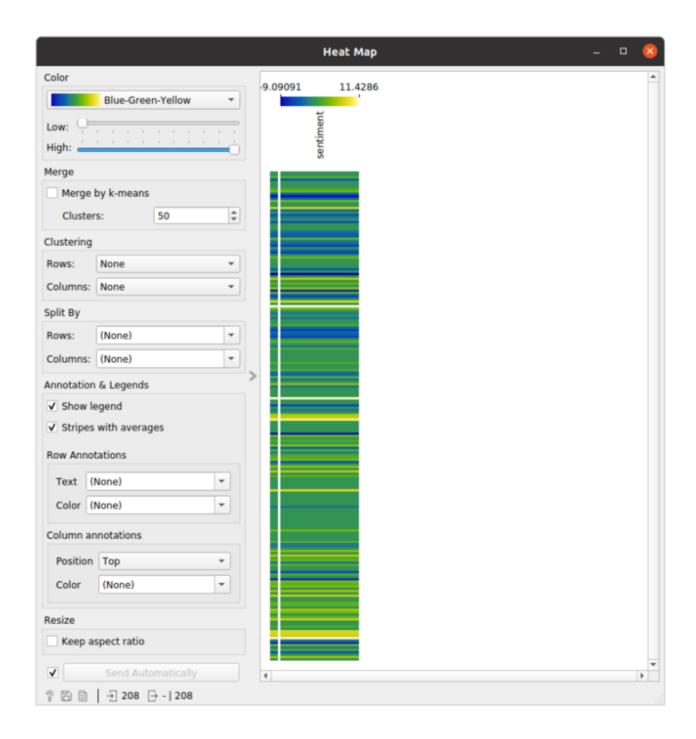
Visualisasi secara keseluruhan dari analisa sentimen di atas dapat ditampilkan menggunakan widget "Heat Map". Silahkan tambahkan widget ini yang diambil dari dock "Visualize".



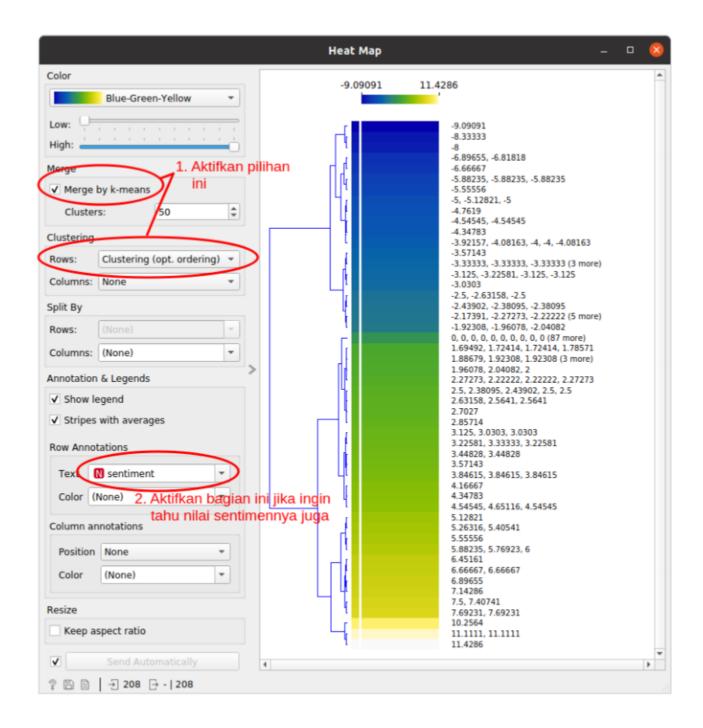
Jika pada waktu menambahkan widget "Heat Map" kemudian muncul ikon tanda seru di atas widget tersebut, itu berarti widget tersebut tidak memperoleh data apa-apa dari widget sebelumnya. Jika hal ini terjadi, lakukan konfigurasi dengan *double-click* pada garis yang menghubungkan widget "Corpus Viewer" dengan "Heat Map" tersebut lalu ubah sebagai berikut:



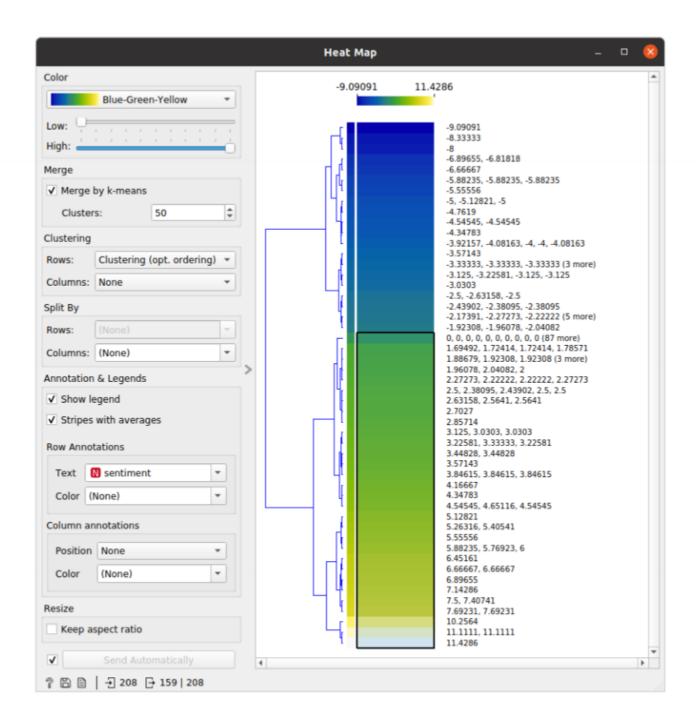
Silahkan double-click widget "Heat Map" untuk melihat bagaimana distribusi sentimen dari cuitan tentang #kripto terjadi. Akan diberikan tampilan sebagai berikut:



Heat Map adalah model visualisasi terhadap variabel tertentu yang nilainya dinyatakan dalam kode warna. Untuk contoh Heat Map diatas, digunakan warna degradasi dari biru menjadi kuning/putih, dimana warna biru menyatakan nilai yang lebih rendah dibanding warna hijau atau bahkan kuning/putih. Agar visualisasinya lebih informatif, kita bisa menambahkan fungsi klasterisasi dengan mengaktifkan beberapa pilihan seperti berikut.



Teknik klastering yang dipakai dalam widget "Heat Map" di atas adalah teknik "Hierarchical Clustering". Dengan memilih/mengaktifkan hirarki yang sesuai, kita bisa melihat bahwa sebagian besar sentimen terkait #kripto di masa itu lebih condong positif atau netral. Berikut adalah visualisasi dari hasil klastering tersebut.



Orange juga menyediakan sarana pengunduh data dari tabel seperti di bawah ini :



Data yang telah diunduh dapat diolah lebih lanjut dengan perangkat olah data lainnya sesuai kebutuhan.

### Daftar Pustaka

https://orange datamining.com/

 $\underline{https://orange3\text{-}text.readthedocs.io/en/latest/widgets/sentimentanalysis.html}$