* 05 / 02 / 2023 – Minggu

1. Mempersiapkan lingkungan / environtment untuk projek, baik dari persiapan docker untuk scrapping twitter (sc : <https://medium.com/@helmisatria/cara-mendapatkan-data-twitter-di-2023-44aaefb616ea> ). Sampai ke melakukan percobaan scrapping data. (pending – docker error). Hari ini malah nonton the art of learning data science by data professor. (Docker bermasalah, dan gatau sebabnya. Hal yang sama juga terjadi pada teknik menggunakan api, kita perlu cara lain untuk scrapping twitter).

* 06 / 02 / 2023 – Senin

Tidak ada perkembangan signifikan mengenai proses projek, karena kemarin hanya belajar sebentar mengenai data science kurikulum. Selain itu, kemarin full kuliah.

* 07 / 02 / 2023 – Selasa
* Melakukan beberapa brainstorm ide mengenai projek / masalah apa yang harus diselesaikan, seperti sebagai berikut :

1. Melakukan sentiment analisis terhadap twit region indonesia, menggunakan dataset twitter. Untuk proses scraping, mungkin perlu pencarian lebih lanjut. Dikarenakan 2 pendekatan (docker, dan api) masih bermasalah. Kita perlu mencari cara scraping lainnya.
2. Terbesit ide untuk menyelesaikan beberapa permasalahan pangan di Indonesia. Selain pangan merupakan komoditas penting manusia, saya rasa juga perlu proses automasi untuk memprediksi berbagai hal di sektor pangan. Hal ini berguna untuk meningkatkan efisiensi dan kinerja pada sektor pangan, agar masyarakat dan bangsa dapat menerima keuntungan lebih. (Untuk detil mengenai permasalahan yang lebih detil mungkin perlu analisis dan perumusan lebih lanjut). Berbicara mengenai permasalahan pangan, kita dapat merumuskan masalah dari dasar terlebih dahulu, sebagai contohnya yaitu dari sisi petani. Berdasarkan mahasiswa pertanian (Ringga Wiyanto a.k.a teman 1 kos) permasalahan pangan yang ada di Indonesia terdiri dari beberapa hal seperti:
   1. Mahalnya harga pupuk, disertai murahnya harga jual dari petani.
   2. Banyaknya proses yang perlu dilakukan oleh petani dalam proses penjualan untuk sampai ke konsumen.
   3. Kurangnya ilmu dan inovasi petani dalam bertani, sehingga menimbulkan inefisiensi pengerjaan petani.

* 08 / 02 / 2023 – Rabu

Mendapati beberapa rumusan permasalahan seperti berikut :

* Indonesia kekurangan guru yang berkualitas sekaligus kurang baiknya perlakuan guru di Indonesia. Buatlah sebuah model yang dapat (?).
* Masalah mindset masyarakat dan sumber ilmu yang didapati atau disembahkan lingkungan Indonesia sangat buruk. Anak-anak dididik untuk menjadi bodoh, dan tidak memiliki kecerdasar berpikir. Proses dasar berpikir manusia didasari oleh apa yang dia amati, dapati, dan rasakan oleh apa yang lingkungan berikan. Jadi, buat sebuah analisis permasalahan mindset di Indonesia, cari sebuah dataset yang dapat mengukur tingkat kecerdasan mindset masyarakat. Berdasarkan hal tersebut, cari segmentasi yang perlu untuk dicerdaskan.
* 09 / 02 / 2023 – Kamis

Membuat sebuah fungsi yang dapat scraping twit dari twitter tanpa perlu API, menggunakan twint. Akan tetapi, kita terkendala masalah error yang belum diketahui sebabnya.

* 10 / 02 / 2023 – Jumat

Berhasil mengidentifikasi permasalahan fungsi scraping sebelumnya, yaitu salah syntax ajha :TT:. Berhasil juga, men-scrapping data twit dari twitter dengan kode pencarian ‘muslim’ bahasa ‘id’ dengan total baris yaitu 10365. Akan tetapi, bermasalah karena terlanjur store di json file. Mungkin kedepannya akan lebih baik jika kita dapat store di csv file.

* 11 / 02 / 2023 – Sabtu

Pada awal hari ini, kita menemukan sebuah project yang menarik yang diambil dari sebuah medium, dan dapat kita jadikan acuan kedepannya. Dengan project itu juga, kita mendapat sebuah program yang dapat scraping html web statis menggunakan salah satu library python ‘*Beautifoul Soup’.* Dengan hal tersebut, kita dapat sejumlah file url dari <https://www.goodreads.com> yaitu website buku. Akan tetapi, scraping yang terjadi sangat lama ☹. Jadi, jangan matikan laptop mu.

Selain beberapa hal tersebut, kita juga telah membuat beberapa fungsi yang dapat digunakan untuk membersihkan data sesuai dengan kebutuhan kita. Data yang akan kita gunakan juga telah bertambah. Besok, lanjutkan bersihkan data, mungkin cek ricek lagi kolom yang dapat diolah. Kemudian, baru kita menganalisis tiap baris. Fungsi yang dibuat antara lain, fungsi en\_only yaitu seleksi language dataset hanya menggunakan bahasa inggris. Selain itu, kita juga membuat fungsi yang membuang beberapa kolom (lengkapnya dapat diliat pada program)

* 12 / 02 / 2023 – Minggu

Beberapa hal yang mungkin akan kita lakukan hari ini :

* + 1. Membuat fungsi yang menambahkan kolom berdasarkan kolom “created\_at”, yang kite perlukan adalah daerah tempat dibuanya (region) >>
    2. Membuat fungsi yang dapat mereturn berapa user yang punya replies, retweets, dan likes terbanyak >>
    3. Coba pikirkan lagi mengenai tujuan awal kita mencari dataset ini adalah untuk apa. Masalah apa yang ingin kita highlight.
    4. Membuat fungsi yang dapat mereturn word counts
* 13 / 02 / 2023 – Senin

Awali pagi dengan stuck, kita menghadapi masalah dengan “id”. Hal ini dikarenakan id yang ada di database terlalu besar, sehingga menggunakan format “e”. Hal ini menyebabkan akses terhadap id tersebut sulit dilakukan dari luar database. Hal ini bertambah masalah dengan tidak tahunya keyword yang tepat untuk mencari solusi dari internet. (1)

Jadi, untuk saat ini mari kita fokus terhadap fungsi words count terlebih dahulu 😊. Asd

* 15 / 02 / 2023 – Selasa

Menambah fungsi wordscount sekalian menggunakan stopwords untuk memilah penggunaan kata yang dibutuhkan. Melakukan visualisasi terhadap data yang sering dikeluarkan (tidak sesuai ekspektasi == kecewa).

Mungkin besok, perlu untuk menambah sebuah fungsi sentiment analysis untuk menganalisis sentiment semua tweet dan mencari sebuah insight yang dibutuhkan. Fokus utama saat ini adalah kembali brainstorming dan mencari tujuan dari projek.