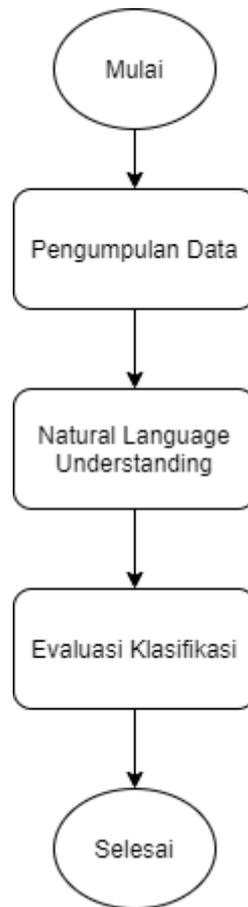


BAB III

METODOLOGI PENELITIAN

3.1 Kerangka Berpikir

Dengan bantuan kemajuan teknologi yang ada, dapat dilakukan serangkaian percobaan dengan hanya menggunakan media sosial sebagai bahan untuk mengetahui emosi seorang pengguna media sosial berdasarkan apa yang ditulis di media sosial. Dalam penelitian ini, sosial media yang digunakan adalah *Twitter* seperti yang telah dibahas sebelumnya pada Bab 1. Penelitian ini akan membahas secara lebih spesifik tentang analisis dan klasifikasi emosi seseorang dalam membuat *tweets* dalam menggunakan sosial media *Twitter*, yang dapat ditunjukkan pada gambar 3.1.



Gambar 3. 1 Kerangka Berpikir

3.2 Metodologi Penelitian

Metodologi yang dipakai dalam penelitian ini adalah sebagai berikut:

1. Pengumpulan data
2. Mengklasifikasikan data menggunakan metode *natural language understanding machine learning*
3. Mengevaluasi klasifikasi

3.2.1 Pengumpulan Data

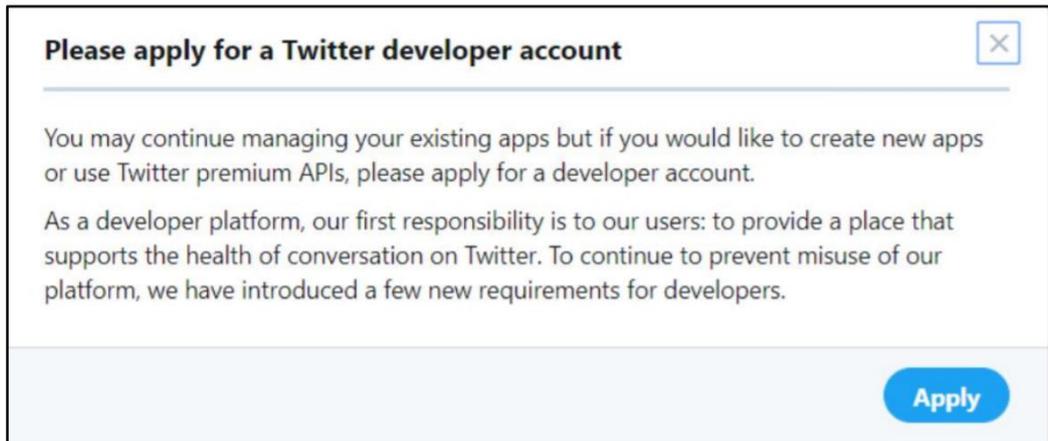
Pengumpulan data pada tahap ini menggunakan library tweepy dalam bahasa pemrograman python. Alasan menggunakan tweepy adalah karena hanya tweepy yang memungkinkan untuk melakukan interaksi dengan API yang telah disediakan oleh *Twitter*. Mengumpulkan data untuk mendapatkan *username* pada *Twitter* yang nantinya akan digunakan saat dalam membuat *code*. Berikut ini adalah tahapan rinci cara mengumpulkan data dari *Twitter*:

1. Mendaftar untuk akses otentikasi pada pengembang *Twitter* di situs web <https://developer.Twitter.com/en/apps>. Masuk dengan akun *Twitter* yang biasanya digunakan. Maka akan muncul gambar 3.2 yang dibawah ini, klik tombol yang dilingkari merah.



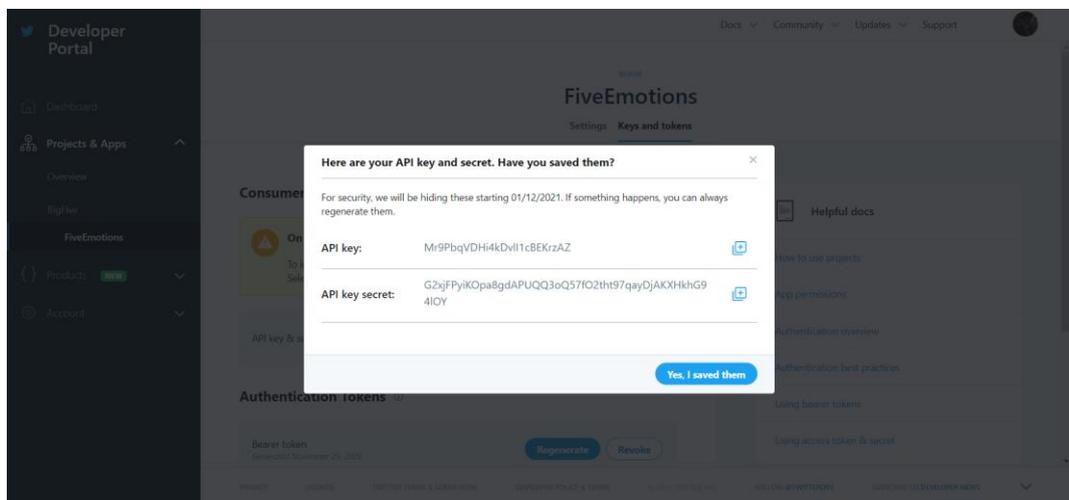
Gambar 3. 2 *Twitter Developer*

2. Kemudian ikuti instruksi di *Twitter* seperti yang tertera pada gambar 3.3 dan menunggu balasan bahwa pendaftaran kami diterima oleh *Twitter* karena pendaftaran ini langsung diperiksa oleh karyawan *Twitter* untuk dilihat maksud dan tujuan penelitian, yang akan memakan waktu agak lama, sekitar 1-3 hari kemudian.



Gambar 3. 3 Pendaftaran Membuat Apps

3. Setelah itu baru bisa membuat aplikasi seperti gambar 3.4 yang ada dibawah ini dan mendapatkan API key dan API key secret yang sangat penting dan berguna nantinya dalam membuat *visual code*.

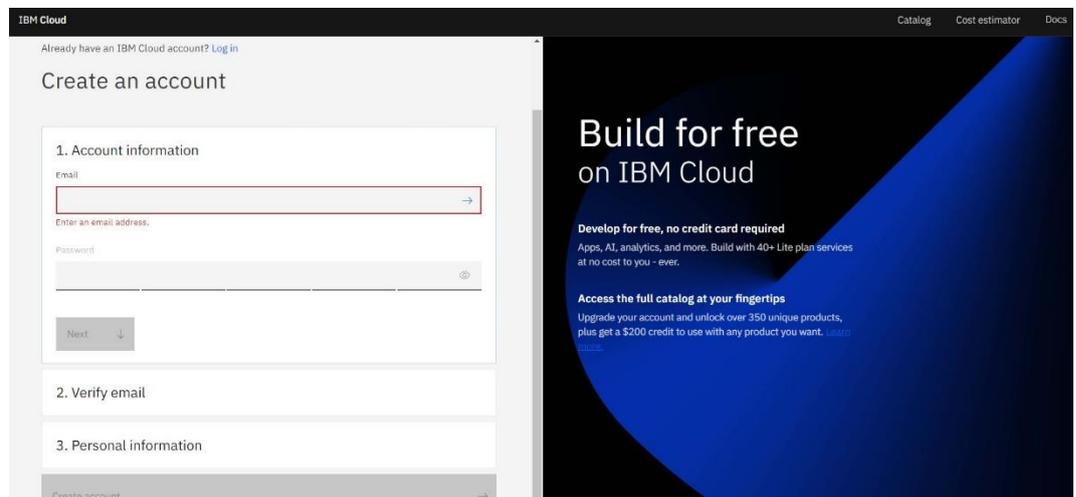


Gambar 3. 4 API Key Twitter

3.2.2 Natural Language Understanding

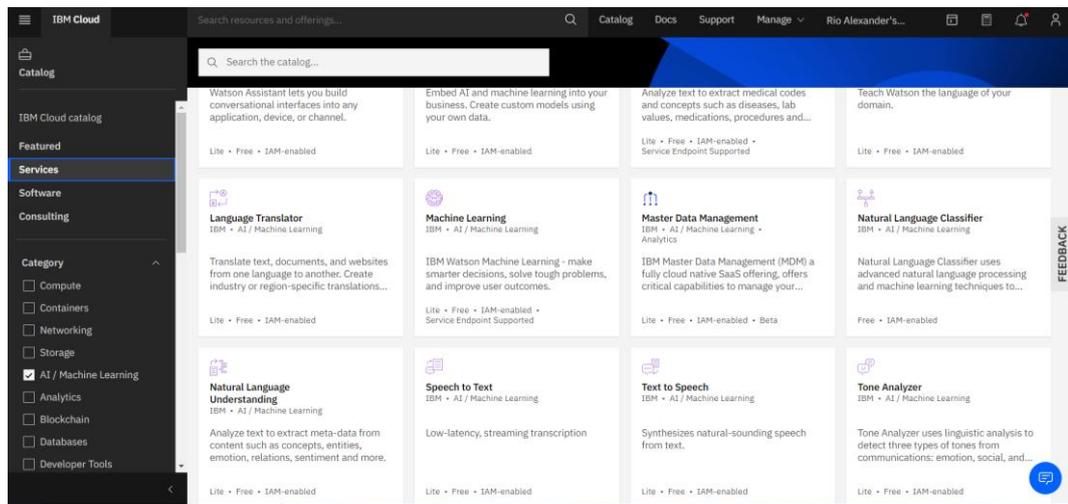
Untuk melakukan analisis dengan metode *Natural Language Understanding*, maka akan digunakan *Function Natural Language Understanding* yang terdapat pada IBM. Alasan menggunakan IBM adalah karena IBM sudah menyediakan berbagai macam *code* yaitu salah satunya adalah *natural language understanding* yang dapat digunakan untuk melakukan klasifikasi ataupun pemrograman. Berikut ini adalah tahapan rinci cara mengambil *code* dari IBM:

1. Mendaftar untuk akses otentikasi pada pengembang IBM di situs *web* <https://cloud.ibm.com/registration> seperti yang terdapat digambar 3.5. Mendaftar dapat menggunakan *email* yang biasanya digunakan, kemudian diverifikasi agar dapat digunakan secepatnya.



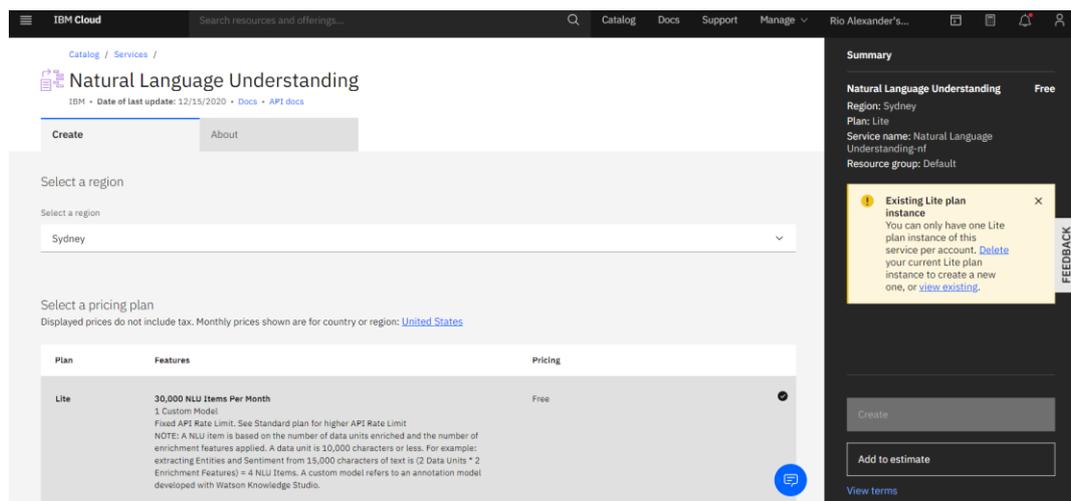
Gambar 3. 5 Daftar IBM

2. Setelah mendaftar, maka terdapat *IBM Cloud catalog* seperti gambar 3.6 yang bisa dipilih untuk diambil codenya yang akan digunakan untuk melakukan klasifikasi dan *code* yang akan digunakan adalah *Natural Language Understanding* dikarenakan NLU merupakan satu-satunya machine learning yang dapat mendeteksi emosi seseorang.



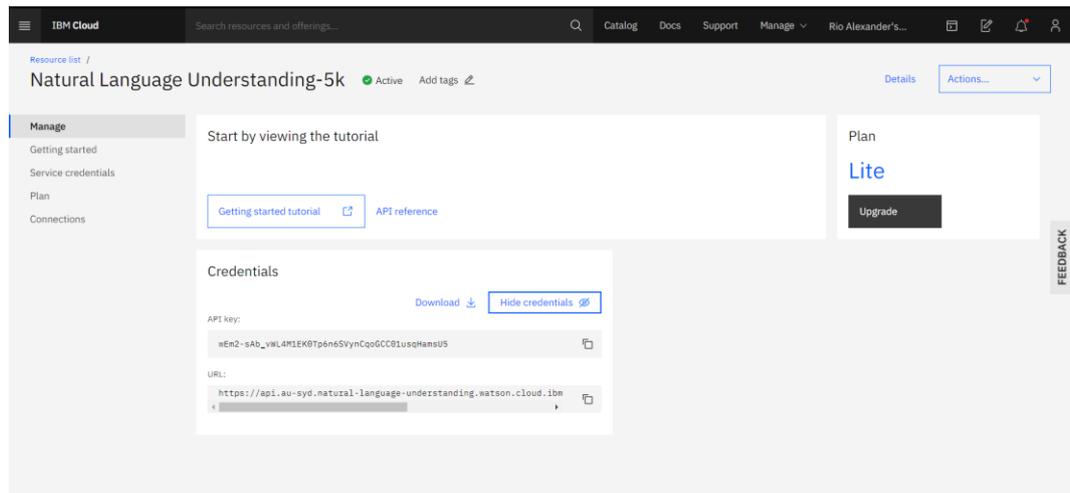
Gambar 3. 6 IBM Cloud Catalog

- Setelah memasuki dalam bagian Natural Language Understanding, maka tinggal mengikuti intruksi dari IBM tersebut dengan memilih *region* dan *pricing plan* yang ingin digunakan, yang dapat dilihat pada gambar 3.7.



Gambar 3. 7 Create NLU

- Setelah *create* NLU, maka dapat ditemukan API key dan URL yang nantinya akan digunakan untuk melakukan klasifikasi pada penelitian ini yang dapat ditemukan pada gambar 3.8 ini.



Gambar 3. 8 API Key IBM

3.2.3 Mengevaluasi Klasifikasi

Setelah itu akan dilakukan proses analisis dan visualisasi data di dalam aplikasi *visual studio code* berupa *bar plot* yang memberikan informasi tentang hasil analisis yang diklasifikasi *five emotions* seseorang, antara lain, perbandingan antara *sadness*, *joy*, *fear*, *disgust*, dan *anger*. Untuk dapat mengevaluasi klasifikasi data, saya menggunakan *library matplotlib* yang merupakan *library Python* yang fokus pada visualisasi data, seperti membuat plot grafik. Visualisasi *matplotlib* merupakan gambar grafis dengan satu sumbu atau lebih. Setiap sumbu memiliki sumbu horizontal (x) dan sumbu vertikal (y) dan data yang direpresentasikan menjadi warna dan *glyphs* seperti *marker* (contohnya bentuk lingkaran) atau *lines* (garis) atau *polygon*. Untuk menggunakan *matplotlib*, lakukan impor terlebih dahulu *library*

matplotlib.pyplot as *plt*. Penggunaan *as* disini, artinya kita menggantikan pemanggilan fungsi *pyplot* pada *matplotlib* dengan *prefix plt* untuk proses berikutnya. Disini terdapat *magic command %matplotlib inline*, untuk pengaturan pada *backend matplotlib* sehingga setiap grafik dapat ditampilkan secara *inline*, yang akan ditampilkan langsung di *visual code studio*.

3.3 Problem Solving

Pada penelitian ini, machine learning yang digunakan untuk menjadi problem solving yaitu *natural language understanding* (NLU) yang merupakan sub-bagian *natural language processing* (NLP), setelah komputer memahami arti dari dokumen / kalimat (tugas klasifikasi dokumen), komputer akan mengeluarkan informasi dari hasil sebelumnya, seperti hasil klasifikasi, entitas, tanda pos, dan lain lain. Informasi ini dapat diproses ulang oleh *engineer*. Salah satu alasan memilih untuk menggunakan NLU adalah karena NLU merupakan satu-satunya *Machine Learning* pada IBM yang dapat mengklasifikasi emosi seseorang melewati kata-kata. Selain itu terdapat perbandingan antara *natural language understanding* (NLU) dan *natural language generation* (NLG). Berikut pada table 3.1 menunjukkan perbedaan *natural language understanding* (NLU) dan *natural language generation* (NLG) berdasarkan kegunaannya.

Tabel 3. 1 Perbandingan NLU dan NLG

<i>Natural Language Understanding</i> (NLU)	<i>Natural Language Generation</i> (NLG)

<p>Lexical Analysis yang mengidentifikasi dan menganalisis struktur kata dan membagi seluruh bagian teks menjadi paragraf, kalimat, dan kata.</p>	<p>Discourse Generation adalah proses dimana input adalah tujuan komunikasi dan output dapat berupa wacana sering kali dalam bentuk <i>content tree</i>.</p>
<p>Syntactical Analysis yang menganalisis kata-kata dalam kalimat untuk tata bahasa dan menyusun kata-kata dengan cara yang menunjukkan hubungan antar kata.</p>	<p>Sentence Planning melibatkan realisasi permukaan atau linierisasi menurut tata bahasa</p>
<p>Semantic Analysis yang menarik makna yang tepat atau makna kamus dari teks.</p>	<p>Lexical Choice melibatkan pemilihan kata konten (kata benda, kata kerja, kata sifat, dan kata keterangan) dalam teks yang dihasilkan.</p>
<p>Discourse Integration yang mengartikan kalimat apa pun bergantung pada arti kalimat tepat sebelumnya.</p>	<p>Sentence Structuring adalah proses pembuatan teks kalimat, yang harus benar sesuai dengan aturan sintaks</p>
<p>Pragmatic Analysis berguna untuk mengartikan ulang tentang apa yang dimaksud.</p>	<p>Morphological Generation merupakan penataan akhir yang mungkin melibatkan koreksi seperti perbedaan tegang atau perbedaan gender dari suatu bahasa dalam konteks entitas</p>

3.4 Teknik Analisis Data

Alat yang digunakan dalam penelitian ini adalah bahasa pemrograman python dan alasan menggunakan python adalah karena python memiliki *testing framework* yang dibutuhkan pada penelitian saya, yang meliputi unit test, nose, doctest, dan pytest, yang digunakan untuk menguji dan memeriksa secara menyeluruh setiap unit codenya dan untuk menjalankan code pengujian yang ditulis dengan formatnya. Selain python terdapat alat analisis yaitu R. Berikut perbandingan antara Python dan R berdasarkan kegunaannya yang terdapat pada tabel 3.2.

Tabel 3. 2 Perbandingan Python dan R

Python	R
<i>Python</i> memiliki <i>library package</i> bernama <i>Matplotlib</i>	Visualisasi data sangat unggul, karena R memiliki <i>library package</i>
Pemrosesan data dan <i>data mining</i>	Analisis data dan perhitungan statistik
Lebih diunggulkan dalam analisis yang berulang	Lebih diunggulkan dalam analisis teks misalnya, membagi paragraf menjadi frasa dan kata-kata
<i>Python</i> sangat baik dalam prediksi akurasi dan <i>machine learning</i> .	R adalah juara di bidang ilmu data karena latar belakang statistika.
Digunakan lebih luas di perusahaan dan organisasi	Lebih banyak digunakan dalam dunia pendidikan