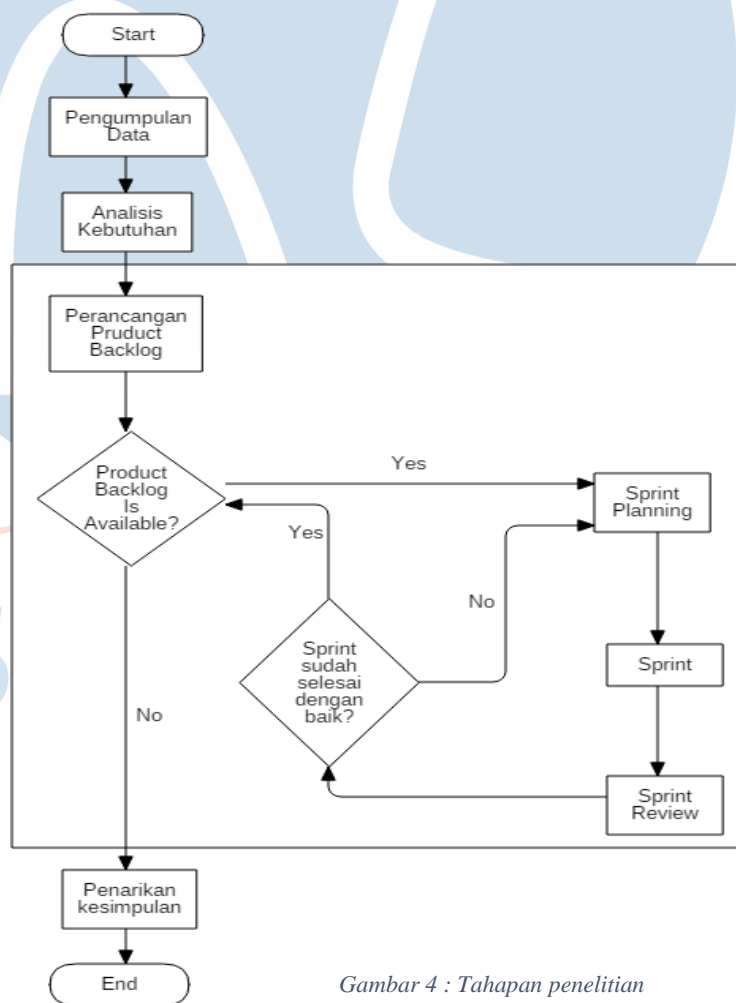


BAB III METODE PENELITIAN

3.1 Tahapan Penelitian

Metode penelitian yang dilakukan pada penelitian ini bersifat kuantitatif. Proses pengambilan data dilakukan dengan menggunakan metode scrum. Berikut gambarnya:

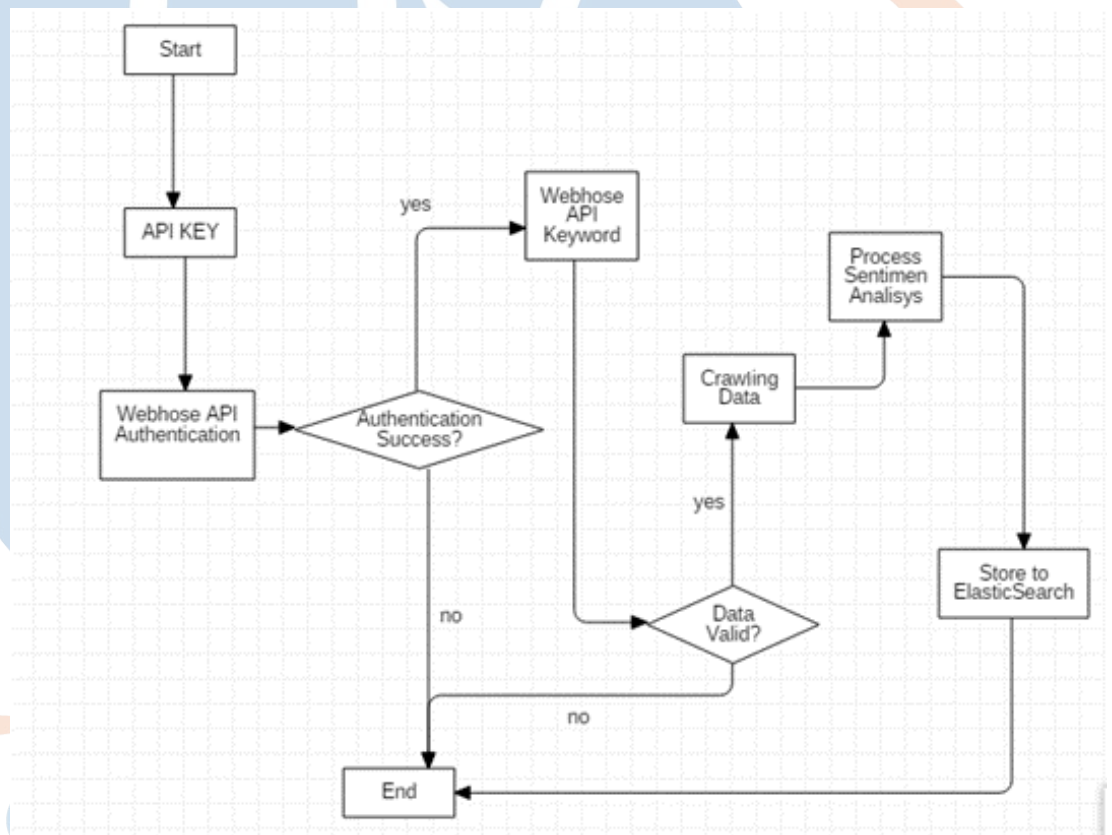


Gambar 4 : Tahapan penelitian

3.2 Rancangan

3.2.1 Pemecahan Masalah

Proses pengumpulan data dilakukan dengan menggunakan Webhose API yang di implementasikan menggunakan bahasa pemrograman Python. Proses yang akan dilakukan Berikut adalah alur proses nya:



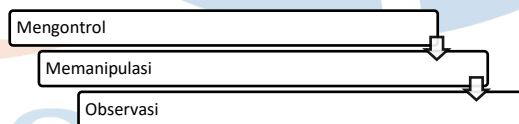
Gambar 5 : Flowchart pemecahan masalah

Pada Proses Crawling data akan di teruskan ke Proses Sentimen Analysis, pada Proses Sentimen Analisis dimulai dari proses tokenisasi yaitu pemisahan rangkaian kalimat menjadi bentuk token atau potongan kata tunggal dan di dalam proses tokenisasi juga menghilangkan tanda baca dan merubah token menjadi huruf

kecil, kemudian dilanjutkan dengan penghapusan stop-words yaitu membuang kalimat yang tidak penting dalam suatu bagian berita, diakhiri dengan proses stemming yaitu mengubah term kembali menjadi bentuk kata dasar. Pada tahap preprocessing juga menghasilkan sebuah inverted index yaitu matriks antara term dan koleksi data berita. Tahap kedua yaitu Inverted index yang telah dihasilkan dari proses pre-processing dibagi menjadi 2 bagian yaitu data latih dan data uji. Dari data latih, dengan menggunakan naïve bayes classifier dilakukan perhitungan untuk menghasilkan nilai probabilitas setiap kelas. Nilai probabilitas setiap kelas yang dihasilkan digunakan untuk melakukan klasifikasi data uji sehingga menghasilkan data uji yang terklasifikasi. Tahap terakhir dari pengklasifikasian data berita ini yaitu mengukur prosentasi ketepatan (akurasi) dalam menentukan klasifikasi dari data uji yang diperoleh dari perhitungan pada tahap kedua. Untuk mengukur akurasi klasifikasi data uji digunakan sebuah confusion matrix yaitu dengan menentukan berapa nilai TP, TN, FP dan FN.

3.2.2 Jenis Penelitian

Metodelogi penelitian yang di lakukan adalah eksperimen, yang merupakan metode inti dari model penelitian yang mengandung pendekatan kuantitatif. Dalam metode eksperimen, peneliti harus melakukan tiga persyaratan yaitu:



Sebagai metode penelitian yang digunakan untuk mencari pengaruh perlakuan tertentu terhadap yang lain dalam kondisi yang terkendalikan menggunakan:

- Meneliti literatur yang berhubungan dengan masalah penelitian
- Mengidentifikasi dan membatasi masalah
- Merumuskan hipotesis
- Menyusun rencana secara lengkap dan operasional, meliputi:
- Menentukan variabel bebas & terikat

- Memilih desain yang digunakan
- Menentukan sampel
- Menyusun alat
- Membuat outline prosedur pengumpulan data
- Merumuskan hipotesis statistik

3.2.3 Metode Pengumpulan Data

Proses pengumpulan data menggunakan Webhose API sebagai source data yang di *crawling* oleh aplikasi *crawler*. langkah awal adalah membuat akun webhose di webhose.io . setelah register akan diberi API Key yang nantinya digunakan sebagai kunci untuk kode authenticator dalam menggunakan Webhose API.

Setelah terhubung parameter yang di kirim adalah keyword dan site. Keyword adalah kata kunci pencarian, Site adalah alamat media online yang menjadi sumber pencarian berita yang di *crawling*.

Hasil *Crawling* di proses analisis sentimen-nya lalu hasil dari pemrosesan tersebut akan di simpan di database.

3.2.4 Lingkungan Pengembangan

Pada tahap pengembangan dilakukan perancangan dan pembuatan aplikasi berdasarkan pemodelan sistem yang sudah di dilakukan. perancangan sistem untuk dilakukan klasifikasi data dengan menggunakan algoritma naive bayes. serta pembuatan menggunakan bahasa pemrograman python menggunakan package nltk untuk dilakukan nya pre-processing text.

Pengumpulan data media online dari beberapa website berita di indonesia, yaitu : tribunnews.com , detik.com, kompas.com, sindonews.com, kumparan.co dan idntimes.com. yang akan di-crawling menggunakan package webhose.io dengan menggunakan bahasa pemrograman python.

BAB IV

ANALISIS DAN PERANCANGAN

Pada Bab ini dijelaskan mengenai proses analisis dan perancangan sistem yang penulis kembangkan dengan menggunakan data dari hasil crawling dengan topic tentang pelantikan presiden dan menteri. Hasil tersebut dibuat menjadi sebuah product backlog yang dijadikan sebagai data untuk dilakukan proses penghitungan sentiment dan trending level pada topic berita tertentu.

4.1 Analisis Sistem

Pada tahap ini penulis melakukan analisis sistem dengan melakukan studi pustaka, dan observasi terhadap hal - hal yang berkaitan dengan perancangan aplikasi melalui internet, jurnal, dan aplikasi serupa.

Di bagian ini penulis menjelaskan beberapa kebutuhan untuk merancang aplikasi yang dibangun dari sisi database yang digunakan menggunakan elasticsearch, proses preprocessing text, sentiment analysis dan penerapan algoritma naïve bayes. fitur yang biasa digunakan dalam pendekatan dengan metode SCRUM. Selain product backlog, penulis juga merancang flowchart yang menggambarkan alur proses mulai dari pengambilan data, preprocessing text, dan sentiment analysis.

4.1.1 Product Backlog

Menghasilkan aplikasi yang berfungsi untuk mengumpulkan berita tertentu sesuai dengan permintaan berupa keyword dan alamat website yang sudah ditentukan serta mengolah berita dengan cara memisah perkalimat, tokenizer dan stemming terhadap setiap kata lalu melakukan tersebut sehingga menghasilkan suatu nilai sentimen terhadap keyword yang diset dari berbagai alamat website yang sudah ditentukan sebelumnya dengan mengimplementasikan algoritma naïve bayes untuk mengetahui kebenaran dari hasil sentimen

4.1.2 User Story

Table 4: User Story

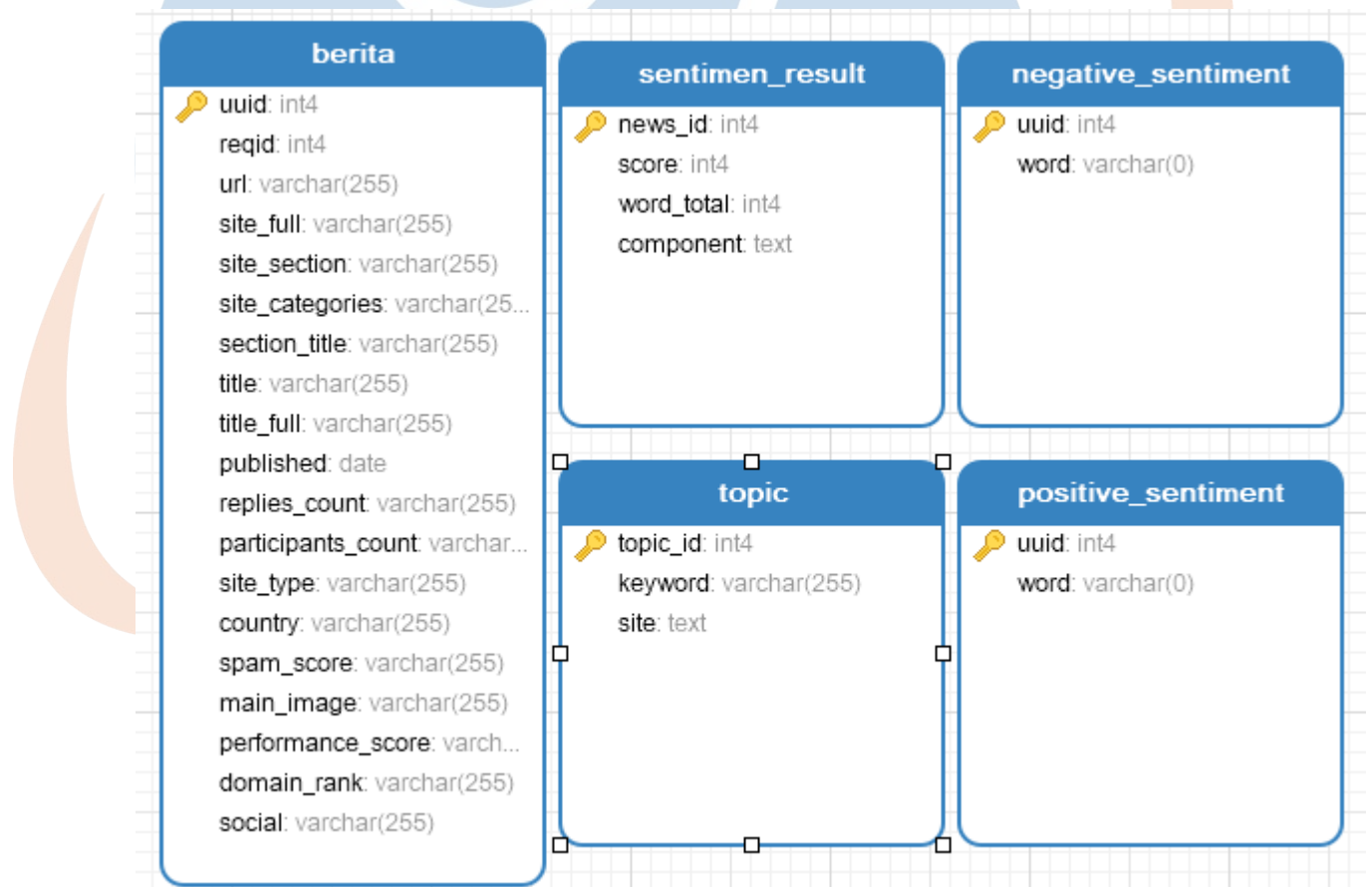
No	Sebagai	Saya Ingin	Sehingga
1	User	Mencari berita berdasarkan topik	Dapat melihat hasil crawling berdasarkan topik yang dicari
2	User	Mencari berita dari sumber tertentu	Dapat melihat hasil crawling berdasarkan sumber tertentu
3	User	Melihat sentimen dari berita yang telah dicrawling	Dapat melihat hasil sentimen dari berita yang telah dicrawling
4	User	Melihat trending rate dari berita yang dicari berdasarkan topik	Dapat melihat seberapa banyak topik tersebut di posting dari berbagai sumber

STT - NF

4.2 Perancangan Sistem

4.2.1 Data Model

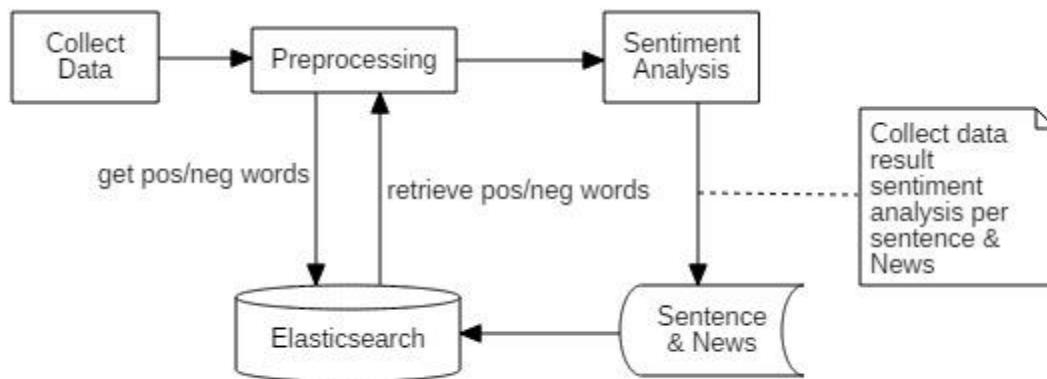
Entity Relationship Diagram (ERD) merupakan salah satu bentuk pemodelan basis data yang sering digunakan dalam pengembangan sistem informasi. Diagram hubungan entitas (ERD) menunjukkan hubungan dari entitas set disimpan dalam database. Entitas dalam konteks ini adalah komponen data. Dengan kata lain, diagram ER menggambarkan struktur logis dari database.



Gambar 6 : Data Model

Gambar diatas adalah erd yang dibuat untuk menyimpan data hasil crawling dan sentiment words yang sudah diload kedalam elasticsearch menggunakan metode bulk insert

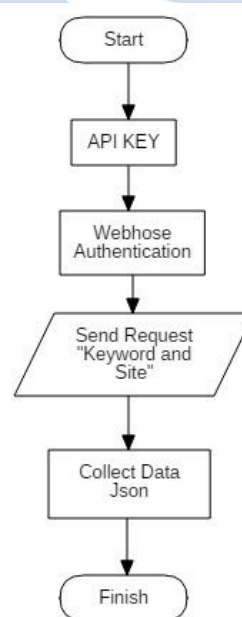
4.2.2 Rancangan Alur Sistem



Gambar 7 : Flowchart keseluruhan proses crawling dan pemrosesan sentimen

Pada gambar diatas adalah gambaran keseluruhan alur system aplikasi, yang mana akan dijabarkan satu persatu dari proses tersebut dengan flowchart dibawah ini.

1. Flowchart Proses Pengambilan data

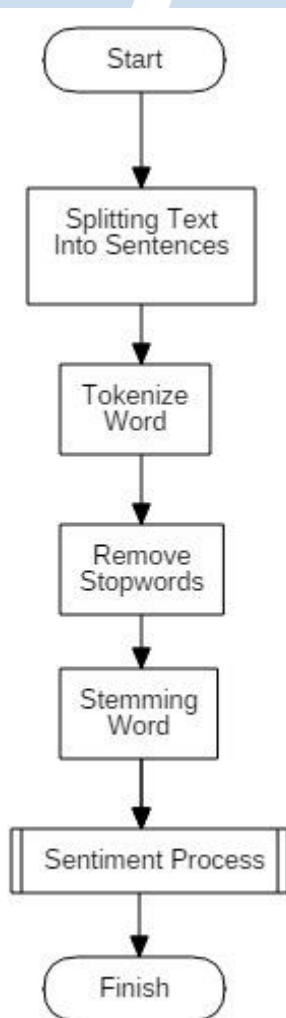


Gambar 8 : Flowchart Proses Crawling Data

Pada gambar dihalaman sebelumnya adalah Flowchart proses crawling data menggunakan Webhose API, membutuhkan API Key untuk melakukan request data dari penyedia data media online yaitu Webhose.

Untuk melakukan request membutuhkan list keyword untuk pencarian berita dan site yang berupa domain media online.

2. Flowchart Proses Preprocessing text

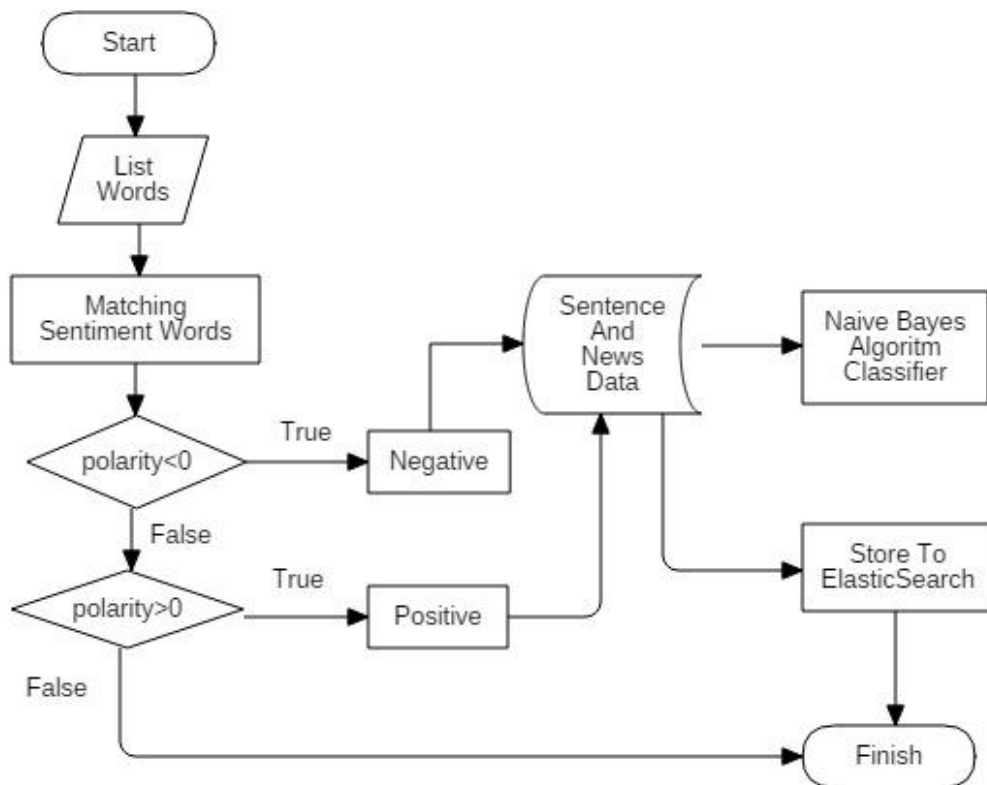


Gambar 9 : Flowchart Proses Preprocessing Text

Pada gambar diatas adalah alur proses dari text processing dari berita yang telah dicrawling, isi konten dari setiap berita yang nantinya akan dipisahkan setiap kalimat nya, setelah kalimat sudah dipisahkan dari konten, akan dilakukan tokenize word yaitu pemisahan kalimat menjadi kata, hasil dari pemisahan ini akan dibersihkan dengan proses penghapusan stopwords yaitu kata umum (common words) yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna.

Setelah melalui proses penghapusan stopwords, setiap kata akan dilakukan proses stemming yaitu pemetaan dan penguraian bentuk dari suatu kata menjadi bentuk kata dasarnya, setelah itu akan dilakukan proses klasifikasi sentiment menggunakan algoritma naïve bayes.

3. Flowchart Proses Sentimen Menggunakan Algoritma Naïve Bayes



Gambar 10 : Flowchart Proses Sentimen Naïve Bayes Classifier

Pada gambar diatas semua hasil dari text processing melalui proses matching sentiment. Yang mengklasifikasikan setiap kata yang termasuk sentiment positif atau negatif, setelah dihitung berapa total positif dan negatif diimplementasikan algoritma naïve bayes classifier untuk menghitung probabilitas sentiment dari berita tersebut.

Proses penghitungan probabilitas menggunakan algoritma naïve bayes diproses berdasarkan polaritas antara kata positif dan kata negatif yang dihitung dengan rumus seperti ini:

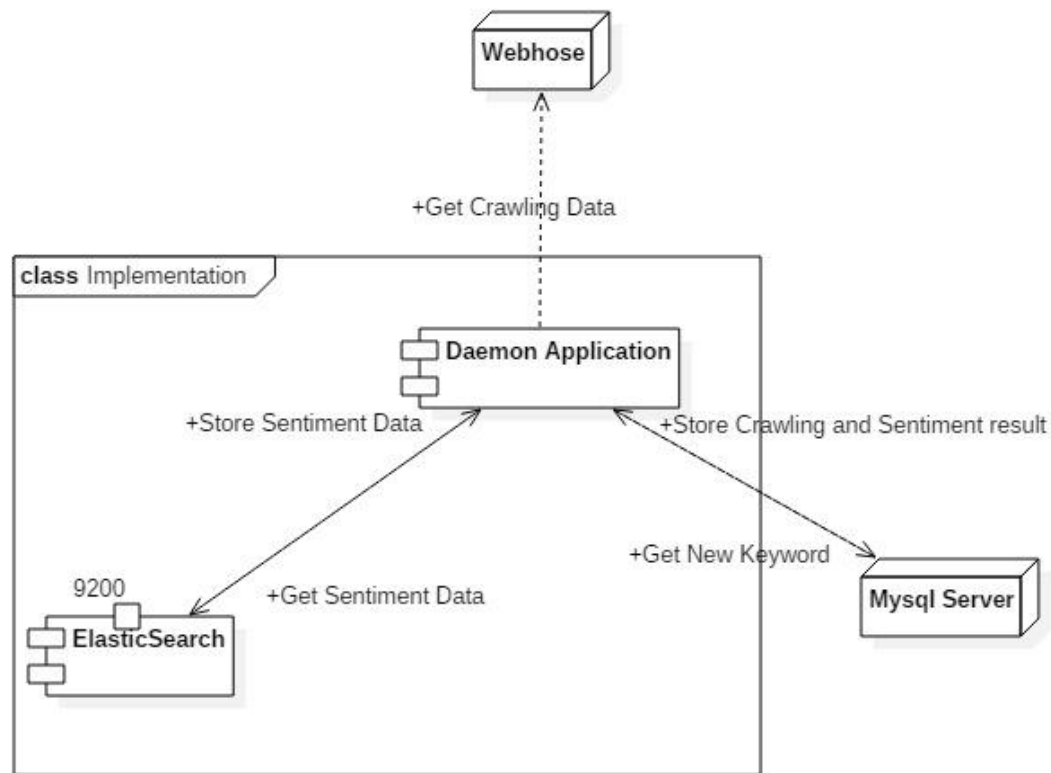
$$\text{Polarity} = \text{positive_word} - \text{negative_word}$$

Jika perhitungan kurang dari 0 maka sentiment adalah negatif, jika perhitungan lebih dari 0 maka hasil sentiment adalah positif dan jika hasilnya adalah 0 maka netral, proses pelabelan data hanya kalimat yang bersentimen positif dan negatif saja.

4.2.3 Deployment

Pada alur proses deployment aplikasi terdiri dari beberapa environment yang saling berhubungan, seperti yang dijabarkan pada proses dibawah ini.

STT - NF



Gambar 11 : Deployment Diagram Aplikasi Crawling

Pada gambar diatas menggambarkan alur deployment aplikasi proses pengambilan data crawling yang dilakukan oleh daemon aplikasi yang dibuat menggunakan bahasa pemrograman python. Juga menggunakan elasticsearch sebagai basis data tidak terstruktur. pengambilan data keyword dan hasil data crawling hasil sentimen akan disimpan ke mysql server. Dibutuhkan internet untuk melakukan proses crawling ke *Webhose.io* dan autentikasi menggunakan API Key yang sudah disediakan setelah membuat akun *Webhose*.

Pada proses deploy ada beberapa tools atau aplikasi yang harus diinstal pada server yang digunakan untuk melakukan proses crawling yaitu:

- Python 3.6
- Mysql 5.7
- Elasticsearch 7.2.0

4.3 Rancangan Pengujian

Perancangan pengujian yang digunakan pada penelitian ini adalah menggunakan metode blackbox testing. Berikut adalah skenario pengujian sistem untuk aplikasi crawling.

4.3.1 Blackbox Testing

No	Aktifitas pengujian	Hasil yang diharapkan	Hasil Pengujian	Kesimpulan
1	Ambil Keyword dari mysql-server	Keyword berhasil diambil dan terkoneksi ke mysql-server		
2	Lakukan Crawling berdasarkan keyword yang sudah diambil	Berita berhasil dicrawling		
3	Lakukan Text preprocessing	Berita berhasil diproses		
4	Lakukan klasifikasi sentiment pada setiap berita	Berita berhasil diproses dan hasilnya sesuai		
5	Simpan hasil crawling dan proses sentiment ke mysql-server	Data tersimpan ke database		

Table 5 : Blackbox Testing