

BAB I

PENDAHULUAN

1.1. Latar Belakang

Menurut (Irwandani, 2016) perkembangan pengguna internet di Indonesia terbilang sangat pesat. Dalam 5 tahun terakhir, jumlah pengguna internet di Indonesia naik sebesar 430 %. Indonesia berada di peringkat kedua setelah Filipina dalam hal pesatnya pertumbuhan pengguna internet di dunia. Hingga November 2015 pengguna internet telah mencapai 88,1 juta orang atau sekitar 34 % dari total jumlah penduduk Indonesia. Pesatnya perkembangan tersebut dipengaruhi oleh semakin populernya penggunaan sosial media di kalangan masyarakat.

Laporan dari *We Are Sosial*, sebuah badan yang meneliti tentang sosial media, menyebutkan bahwa jumlah pengguna internet sampai Januari 2015 adalah 7.210 milyar pengguna dan pengguna sosial media yang aktif berjumlah 2.078 milyar user. Pengguna sosial media ini mencapai 29% dari populasi penduduk dunia. Dari laporan tersebut juga tertulis bahwa di Indonesia sendiri terdapat 28% pengguna aktif sosial media dari populasi total (sekitar 72 juta pengguna). Per Januari 2015, sosial media yang paling banyak digunakan di seluruh dunia adalah *Facebook* (sekitar 1.366 juta pengguna), di ikuti oleh *Qzone* (sosial media di China), *Google+*, *Instagram*, *Twitter*, dan *Tumblr*. Fenomena yang terjadi di

Indonesia agak sedikit berbeda, sosial media yang paling banyak digunakan adalah *Facebook*, *Twitter*, *Google+*, *LinkedIn*, *Instagram*, dan *Pinterest* (Wijanto, 2015). Hal ini ditandai dengan *posting* data menyangkut keseharian, bisnis, status, gambar, bahkan info terupdate lebih mudah ditemukan di kedua sosial media tersebut.

Sosial media kini menjadi tempat sebagai perbincangan publik. Topik pembicaraan dalam sosial yang beragam menjadi hal unik tersendiri untuk diteliti. Dimana dari topik yang mereka bicarakan menghasilkan opini - opini yang sebagian besar juga mengandung opini sentimen. Contohnya topik viral yang sering dibicarakan saat ini adalah *digital payment* yang merupakan hasil analisis kebutuhan perilaku manusia dan menghasilkan inovasi baru berupa sebuah alat pembayaran *digital*. Saat ini banyak jenis *digital payment* yang telah di kenal masyarakat dua dari mereka adalah *T-cash* dan *Go-pay*, keduanya sama – sama *digital payment* yang memiliki jumlah pengguna yang banyak saat ini. Contohnya saja *t-cash*, dimana *T-cash* telah sukses menjangkau sekitar 25 juta pelanggan di 34 provinsi, serta bisa digunakan di lebih dari 75.000 *merchant outlets* (“Tentang Kami | SEMUA BISA #pakeTCASH,” t.t.). Sama halnya dengan *t-cash*, *go-pay* juga *digital payment* yang memiliki jutaan pengguna yang mengemari aplikasi raksasa dengan vendor *Go-jek* ini.

Dari keduanya lalu muncul pertanyaan siapakah yang lebih baik? untuk menjawab pertanyaan tersebut kita perlu data, ada beberapa cara yang bisa dilakukan untuk mendapatkan data contohnya data opini pengguna yang bisa didapatkan dari media sosial, dengan melihat tren sosial media dan melihat

ketersediaan data yang lebih dari cukup, dalam penelitian ini menentukan sosial media yang menjadi sumber data adalah *Instagram*. Dimana selain jumlah *followers* akun objek yang banyak dibandingkan sosial media yang lainnya pengguna juga lebih aktif berkomentar (tidak pasif). Komentar – komentar tersebut diambil dan dikumpulkan dengan metode *Web Scraping* (mengeruk informasi dari web dengan fokus tertentu dalam kasus penelitian ini adalah komentar pengguna) dengan *tools Jarvee* yang memanfaatkan satu akun pengguna *instagram* sebagai media sarana pengumpulan data dengan metode *web scraping*. Selain itu komentar / opini tersebut juga disebut dengan perspektif, perspektif bermacam – macam akan tetapi beberapa diantaranya mengandung sentimen perspektif. Perlunya analisis untuk menarik kesimpulan adalah untuk melihat perbandingannya dan bisa dijadikan bahan untuk pengembangan teknologi informasi serupa. Hal itu perlu di lakukan untuk penelitian ini guna melihat perbandingan opini pengguna terhadap kedua *digital payment* tersebut melalui sosial media yaitu *instagram*. Maka dari itu untuk mengukurnya dapat dilakukan dengan analisis sentimen atau *sentiment analysis*.

Analisis Sentimen atau *Sentiment analysis* adalah riset komputasional dari opini, sentimen dan emosi yang diekpresikan secara tekstual (Zulfa & Winarko, 2017). Akan tetapi data bisa saja sangat banyak oleh karena itulah penelitian ini memanfaatkan *text mining* yang tujuannya untuk dapat mengekstraksi text sentimen dan juga menentukan *trending topik*. Dalam melakukan *sentiment analyst* sendiri penelitian ini menggunakan *Orange Data Mining. Tools* ini adalah mesin belajar *open source* dan *visualisasi* data untuk pemula dan ahli. Alur kerja analisis data yang *interaktif* dengan *toolbox* yang banyak (“Orange – Data Mining

Fruitful & Fun,” t.t.). *Orange Data Mining* bekerja dengan bahasa pemrograman *python* (“Orange Data Mining Library — Orange Data Mining Library 3 documentation,” t.t.). Pada dasarnya target analisis sentimen adalah untuk menemukan pendapat, umpan balik atau ulasan, dan kemudian mengidentifikasi sentimen yang ingin mereka ekspresikan dan kemudian mengklasifikasikan *polarity* disana yaitu *positif*, *negatif* atau *neutral* (Wu., dkk., (2011) dalam (Anshuman dkk., 2017). Maka dari itu untuk dapat mendapatkan *class sentimen* tersebut, penelitian ini menggunakan metode analisis *vader: Lexicon- and Rule-Based Sentiment Analysis*. Metode ini menentukan class sentimen berdasarkan *lexicon (a library)*, dengan metode analisis *vader*, analisis sentimen akan menghasilkan *Output score* untuk beberapa kategori (*negatif*, *positif*, *neutral*) dan menambahkan *score* sentimen total yang disebut *compound (combined score)* dimana *negative score* untuk *negative sentiment*, *positive* untuk *positive*, 0 adalah *neutral* (“Sentiment Analysis — Orange3 Text Mining documentation,” t.t.). Hal tersebut untuk mengetahui perbandingan analisis sentimen pengguna *T-cash* dan *Go-pay* disosial media.

Vader merupakan metode analisis yang menghasilkan *class sentimen* berdasarkan *lexicon* dengan *source* bahasa *en (english)*, bahasa dari komentar pengguna *T-cash* dan *Go-pay* sendiri adalah bahasa indonesia karena kebanyakan penggunanya adalah warga negara indonesia. Dari penelitian sebelumnya beberapa hal dilakukan untuk dapat mengkategorikan *class sentimen*, seperti pada penelitian (Buntoro, 2017), dimana mengkategorikan *class sentimen* dengan metode *lexion-based* dengan bantuan ahli bahasa indonesia. Hal ini tidak efektif untuk jenis penelitian dengan jumlah data yang besar selain biaya yang mahal,

memerlukan waktu yang lama pula pada proses ini. Selain itu bahasa pemrograman *python* juga sering digunakan untuk mengkategorikan *class sentimen*. Metode analisis *vader* merupakan *package* bahasa pemrograman *python* dari fitur NLTK (*Natural Language Toolkit*) bersamaan dengan metode *liu hu* dan *sentimenalyzer tools* yang dapat digunakan untuk pengkategorian *class sentiment* pada *Orange Data Mining* dengan *source* bahasa *lexicon vader* sendiri yaitu *en (english)* (“*nltk.sentiment package — NLTK 3.4 documentation,*” t.t.-a). Maka dari itu dalam proses mengumpulkan data dengan teknik *Information Scraping* pada penelitian ini memanfaatkan MT (*machine translating*) menggunakan *function GoogleTranslate()* dari *Google drive spreadsheet* dengan *filtering* ejaan terlebih dahulu barulah di *translate* guna mendapatkan data dengan *source* bahasa *en (english)* untuk menunjang proses penelitian berbasis *machine learning* menggunakan *Orange Data Mining*.

Penelitian ini mencoba melihat perbandingan dari seberapa positif, negatif, dan netral dari hasil analisis sentimen menggunakan *Orange Data Mining*. Data yang diambil dalam kurun waktu beberapa bulan postingan terakhir akan dianalisis dan direkapulasi hasilnya lalu barulah hasil rekapulasi dibandingkan. Maka dari itu untuk melihat perbandingan analisis sentimen dari *T-cash* dan *Go-pay* dengan data yang valid dan bersifat publik di sosial media serta pengambilan data yang ekonomis dengan teknik *web scraping*, dan di bantu *tools* analisis sentimen yang mendukung proses penelitian dengan metode analisis yang efektif yaitu *vader*, hal ini untuk meminimalisirkan proses penelitian *non-teknologi* maka dari itu penulis mengajukan penelitian dengan judul **Perbandingan Analisis**

Sentimen Terhadap Digital Payment “T-Cash Dan Go-Pay” Menggunakan Orange Data Mining Di Sosial Media”.

1.2. Perumusan Masalah

Dari uraian latar belakang diatas maka rumusan masalah dari penelitian ini adalah: Bagaimanakah perbandingan hasil analisis sentimen mengenai *digital payment T-cash* dan *Go-pay* di sosial media?

1.3. Batasan Masalah

Dalam penelitian ini memiliki beberapa batasan – batasan permasalahan yang akan diteliti berikut ini batasan masalahnya:

1. Penelitian ini menggunakan sumber data dari media sosial yaitu Instagram dengan menganalisis opini/perspektif sentimen pengguna mengenai digital payment T-cash dan Go-pay menggunakan tools Orange Data Mining.
2. Tools yang di gunakan adalah Orange Data Mining dengan metode analisis sentimen yaitu Vader.

1.4. Tujuan dan Manfaat Penelitian

Dalam melakukan penelitian ini penulis memiliki tujuan penelitian yang mana dapat bermanfaat bagi orang lain. Berikut ini tujuan dan manfaat penelitian ini:

1.4.1. Tujuan

Dalam melakukan penelitian ini penulis memiliki beberapa tujuan diantaranya sebagai berikut:

1. Mengetahui perbandingan antara *digital payment T-cash* dan *Go-pay* melalui analisis sentimen di sosial media.
2. Melakukan analisis sentimen menggunakan metode analisis sentimen *Vader*.

1.4.2. Manfaat

Berikut ini manfaat yang diberikan setelah membaca penelitian ini yaitu diantaranya:

a) Bagi Mahasiswa

1. Melatih mahasiswa untuk bisa memanfaatkan *text mining* dalam mengolah data berupa text dalam jumlah yang banyak dari sosial media.
2. Memberi pengetahuan kepada mahasiswa tentang kegunaan sosial media untuk analisis sentimen
3. Memberi pengetahuan kepada mahasiswa sebagai pengguna *T-cash* atau *Go-pay* tentang perbandingan kedua digital payment tersebut melalui Analisis sentiment.
4. Memberi pengetahuan kepada mahasiswa tentang bagaimana menggunakan *Orange Data Mining* untuk Analisis Sentimen

b) Bagi Ilmu Pengetahuan

1. Diharapkan hasil penelitian ini dapat menjadi kajian lebih lanjut pada penelitian sejenisnya.

1.5. Metodologi Penelitian

1.5.1. Metode Penelitian

Pada dasarnya target analisis sentimen adalah untuk temukan pendapat, umpan balik atau ulasan, dan kemudian mengidentifikasi sentimen yang ingin mereka ekspresikan dan kemudian mengklasifikasikan *polarity* disana yaitu *positif*, *negatif* atau *neutral* (Wu., dkk., (2011) dalam (Anshuman dkk., 2017).

Metode penelitian ini adalah pengabungan metode *Lexicon-based Approach*, metode berbasis *lexicon (a library)* merupakan pendekatan tanpa pengawasan, tetapi dalam hal ini bisa menggunakan sebuah kamus dengan antonim dan sinonim dari kata-kata dan frase opini mereka dengan orientasi sentimen masing-masing (Cambero, 2016) dengan metode *Rule-based Sentiment Analysis* metode ini secara efektif menggunakan algoritma *rule mining* untuk menemukan fitur suatu produk dan untuk mengetahui pendapat yang terkait dengan produk tertentu (Karim & Das, 2018). *Vader* merupakan jenis metode analisis sentimen dengan *lexicon* yang berbasis *rule-based sentiment analysis* dengan mengkategorikan *class sentimen (positif, negatif, neutral)* dan menambahkan skor total atau *compound* (“Sentiment Analysis — Orange3 Text Mining documentation,” t.t., hlm. 3). Untuk melakukan analisis sentimen berbasis *rule-based sentiment analysis* dapat mengikuti kerangka kerja berikut (Kundi dkk., 2014) di sesuaikan dengan alur kerja penelitian ini:

1. Pengumpulan data (*Scarping web* dari *Instagram*)
2. *Preprocessing Module*
3. *Lexicon Module* (*Vader Lexicon- rule- based sentiment analysis*)
4. Identifikasi *Text-Subjektif* (identifikasi sumber data dengan *vader lexicon* atau mengkategorikan *class sentimen*)
5. *Additional Knowledge Module* (Mempelajari pengetahuan tambahan terkait penelitian. Dalam hal ini mencoba menambah *stopword*, dll dengan *re-analysis* untuk menemukan hasil analisis yang lebih baik).

1.5.2. Metode Pengumpulan Data

Dalam penelitian ini bagian – bagian penelitian akan dilakukan secara *komputerisasi*. Mulai dari pengambilan data dengan teknik *web scraping* hingga analisis sentimen menggunakan *Orange Data Mining*.

1.5.2.1. Web Scraping

Aplikasi *web scraping* (juga disebut *intelligent, automated, atau autonomous agents*) hanya fokus pada cara memperoleh data melalui pengambilan dan ekstraksi data dengan ukuran data yang bervariasi. Pada kasus penelitian ini mencoba mencari informasi data yang hanya berupa komentar (sebagai fokus) dari sebuah akun *instagram* selain itu data yang tidak terkait (komeentar) tidak akan di *scrape* (*korek/keruk*). Metode *web scraping* dalam penelitian ini menggunakan *aplikasi scraping* yaitu *Jarvee* untuk *windows* yang memerlukan akun pengguna *instagram* untuk dapat mengambil informasi (komentar) dari sebuah link posting kemudian diekstrak dalam bentuk file format *.csv.

Search engine mengumpulkan informasi dari *web* melalui program *bot* (*robot*) atau *web crawler* yang secara periodik menelusuri *web* (Josi & Abdillah, 2014). *Web scraping* memiliki sejumlah langkah - langkah (Josi & Abdillah, 2014) sebagai berikut:

1. *Create Scraping Template*: Pembuat program mempelajari dokumen *HTML* dari *website* yang akan diambil informasinya untuk *tag HTML* yang mengapit informasi yang akan diambil,
2. *Explore Site Navigation*: Pembuat program mempelajari teknik navigasi pada *website* yang akan diambil informasinya untuk ditirukan pada aplikasi *web scraper* yang akan dibuat,
3. *Automate Navigation and Extraction*: Berdasarkan informasi yang didapat pada langkah 1 dan 2 di atas, *aplikasi web scraper* dibuat untuk mengotomatisasi pengambilan informasi dari *website* yang ditentukan, dan
4. *Extracted Data and Package History*.

Manfaat dari *web scraping* ialah agar informasi yang dikeruk lebih terfokus sehingga memudahkan dalam melakukan pencarian sesuatu. (Juliasari & Sitompul, 2012) dalam (Josi & Abdillah, 2014).

1.5.3. Metode Pengolahan Data

Sebelum analisis sentimen data masih berupa data mentah hasil *Scraping* dari web dan masih mengandung beberapa simbol, aksen, dll yang akan diproses menggunakan metode pengolahan data *Preprocess text*. Berikut ini metode pengolahan data dalam penelitian ini:

1.5.3.1. *Preprocess Text*

Preprocess Text membagi teks sehingga menjadi unit - unit yang lebih kecil (token), *filtering*, melakukan normalisasi (*stemming*, *lemmatization*), membuat *n-gram* dan *tag* token dengan label *part-of-speech*. Langkah-langkah dalam analisis adalah diterapkan secara berurutan dan dapat diaktifkan atau dinonaktifkan. *Preprocess text* akan melaporan beberapa hal seperti jumlah dokumen (tentang jumlah dokumen di memasukkan), total token (menghitung semua token dalam *corpus* (*kumpulan dokumen*)), dan token yang dilaporkan hanya pada token unik di *corpus* bukan token duplikat (“Preprocess Text — Orange3 Text Mining documentation,” t.t., hlm. 3). Dalam *Preprocess text* data penelitian ini akan melali tahap – tahap berikut:

1. ***Transformation***, Mengubah data input. Meliputi: ***Lowercase***, akan mengubah semua teks menjadi huruf kecil. ***Remove Accents***, akan menghapus semua diakritik / aksen dalam teks: contoh : *naïve* → *naive*. ***Parse html*** akan mendeteksi *tag html* dan menguraikan teks saja .: *<a href..>* Beberapa teks ** → Beberapa teks. ***Remove url*** akan menghapus *url* dari teks. Ini sebuah url *http://orange.biolab.si/*. → Ini sebuah url.

2. ***Tokenization***, Tokenisasi adalah metode memecah teks menjadi komponen yang lebih kecil (kata, kalimat, *bigrams*). Meliputi : ***Word dan Punctuation*** akan membagi text berdasarkan kata dan tetap membiarkan symbol tanda baca (tidak menghilangkannya) contoh: *This Sample*. → *(This),(sample),(.)*. ***Tweet***, yang akan membagi teks dengan model *Twitter pra-trained*, yang memuat *hashtag*, *emoticons* dan simbol khusus lainnya. Contoh : *This*

words. *:-) #simple* → (*This*), (*Words*), (*.*), (*:-)*), (*#simple*). Pada dasarnya *Word* dan *Punctuation* serta *tweet* memiliki kesamaan sifat proses akan tetapi *word* dan *punctuation* menjadi proses utama dalam *tokenization*. *Word* dan *punctuation* juga digunakan untuk analisis tren.

3. **Normalization**, berlaku *stemming* (menguraikan) dan *lemmatization* (memilah) kata-kata. (contoh *I've always loved cats* → *I have always loved cats*). Proses ini akan cenderung menggunakan **wordnet lemmatizer** proses ini mencoba mencocokkan text (sinonim / prediksi text) berdasarkan *databases lexicon* yang besar dari NLTK. Proses ini juga mencoba menganalisis typo dalam text (karena penggunaan MT (*Machine Translate*)).
4. **Filtering**, Pemfilteran menghapus atau menyimpan pilihan kata. Meliputi: **Stopwords** menghapus kata - kata penutup dari teks (misalnya menghapus '*and*', '*or*', '*in*' ...). dapat juga memuat daftar kata-kata sandi sendiri yang disediakan dalam format **.txt* file dengan satu *stopword* per baris. Dengan opsi pilihan bahasa *filtering* dapat digunakan untuk banyak bahasa, bahasa *english* ditetapkan sebagai *default*. **Regex**, menghilangkan kata-kata yang cocok dengan ekspresi reguler ini `\.,|:|;|!|\?|\(|\)|\|\\+|'|" | _'|—|'|\\'|...\\-|_|—|\\$|&|*|#|@|%|_|>|<|\\|\\[|\\]|1|2| 3|4|5|6|7|8|9|10|`. Dan secara *default* diatur untuk menghapus tanda baca. **Most Frekuensi Token** (kata kemunculan terbanyak) dengan *word cloud* pada *Orange Data Mining* pada dasarnya untuk melihat kalimat apa yang sering muncul dalam sebuah dokumen dan menentukan berapa banyak yang akan ditampilkan hal ini menggunakan **Most Frekuensi Token**.

1.5.4. Metode Analisis Data

Data hasil *Preprocess text* selanjutnya akan dianalisis dengan metode analisis sentimen berbasis *rule based sentiment analysis* yaitu *vader*. Proses analisis data bertujuan untuk mendapat kategori *class sentimen* (*positif, negatif, neutral*). Text hasil *Preprocess text* akan di nilai berdasarkan *lexicon* apakah itu *positif, negatif* atau *neutral* dan menambahkan skor total (*compound*).

1.5.4.1. Vader : Lexicon- and Rule-Based Sentiment Analysis

VADER (*for Valence Aware Dictionary for Sentiment Reasoning*), termasuk jenis analisis sentimen yang didasarkan pada *lexicon* (*a library*) kata-kata yang terkait sentimen. Dalam pendekatan ini, masing-masing kata di dalam *lexicon* dinilai apakah itu *positif* atau *negatif*, dan dalam banyak kasus, seberapa jumlah *positif* atau *negatif*. Di bawah ini kutipan dari *VADER Lexicon*, di mana lebih banyak kata *positif* memiliki peringkat *positif* lebih tinggi dan lebih banyak kata *negatif* memiliki peringkat *negatif* yang lebih rendah.

Metode analisis *vader* merupakan *package* bahasa pemrograman *python* dari fitur NLTK (*Natural Language Toolkit*) bersamaan dengan metode *liu hu* dan *sentimenalyzer tools* yang dapat digunakan untuk pengkategorian *class sentimen* pada *Orange Data Mining* dengan *source* bahasa *lexicon* metode *vader* yaitu *en* (*english*) (“*nlk.sentiment package — NLTK 3.4 documentation*,” t.t.-b). *Orange Data Mining* bekerja dengan bahasa pemrograman *python* dan terkoneksi dengan NLTK *server* yang menyediakan banyak *corpora, grammer, trained models*, dll terkait studi data mining. Ketika *Vader* mulai bekerja, beberapa perintah menggunakan bahasa pemrograman *python* akan dikerjakan yang paling

penting adalah ketika *vader* memanggil data *lexicon* dari *server* NLTK untuk menghitung *polarity class sentimen*. Berikut ini perintah *connection lexicon vader* di *Orange Data Mining*:

Kode Program 1.1. Connection Vader Lexicon

```

1 class SentimentIntensityAnalyzer(object):
2     """ Give a sentiment intensity score to sentences."""
3     def __init__(
4         self,
5         lexicon_file="sentiment/vader_lexicon.zip/vader_lexicon/vade
6         r_lexicon.txt" ):
7         self.lexicon_file = nltk.data.load(lexicon_file)
8         self.lexicon = self.make_lex_dict()
```

Perintah tersebut memanggil data *lexicon* dari *server* NLTK yang beralamat di http://www.nltk.org/nltk_data/ dimana file tersebut merupakan kamus besar bahasa en (*english*) dan dilengkapi *score* pada setiap kata. Setiap kata akan diberi intensitas skor berdasarkan *lexicon (databases lexicon NLTK dalam jumlah yang besar)*.

1.5.4.2. Vader Determine (Vader Standardization)

Menurut (Hutto & Gilbert, 2014) pada *vader*, text diberi peringkat pada skala dari “[−4] Sangat Negatif” hingga “[4] Sangat Positif”, dengan penyisihan “[0] Netral”. Skor sentimen sebuah kalimat dihitung dengan menjumlahkan skor sentimen dari setiap kata yang terdaftar dalam *Vader lexicon* dalam kalimat. Pembaca yang berhati-hati mungkin akan memperhatikan bahwa ada kontradiksi: kata-kata individu memiliki skor sentimen antara -4 sampai 4, tetapi nilai sentimen akan dikembalikan dengan nilai antara -1 hingga 1. *Score polarity* setiap class akan dinormalkan antara -1 (paling negatif) dan 1 (paling positif) dengan penyisihan 0 adalah netral. Hasil *score compound* mewakili total skor sentimen

yaitu dimana nilai -1 adalah yang paling negatif dan 1 paling positif (“Sentiment Analysis — Orange3 Text Mining documentation,” t.t., hlm. 3). Berikut ini proses normalisasi dari metode analisis sentimen *vader*:

Kode Program 1.2. Normaliasi Skor Text

```

1 def normalize(score, alpha=15):
2     """
3     Normalize the score to be between -1 and 1 using an alpha that
4     approximates the max expected value
5     """
6     norm_score = score / math.sqrt((score * score) + alpha)
7     return norm_score

```

Atau secara matematika menurut (Hutto & Gilbert, 2014), adalah sebagai berikut:

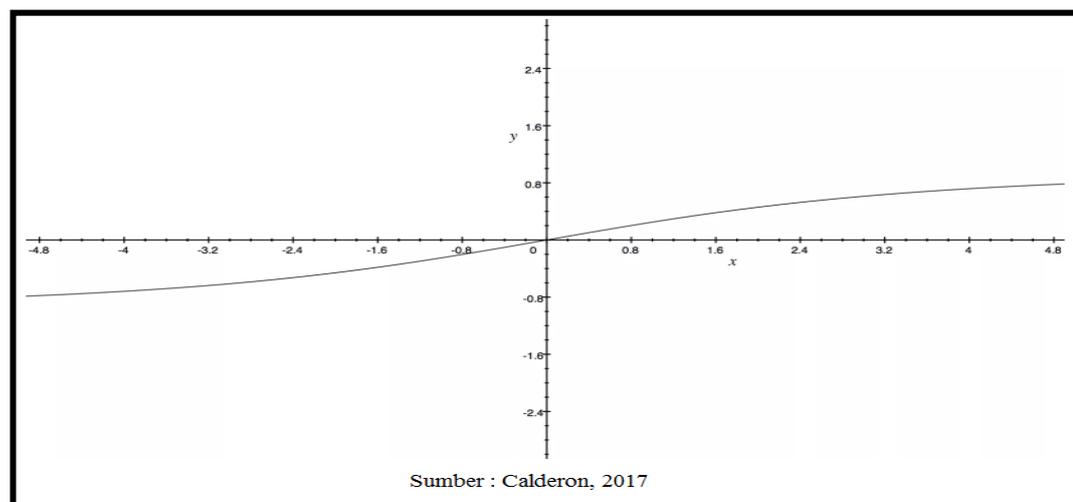
$$norm_{score} = \frac{x}{\sqrt{x^2 + \alpha}}$$

Keterangan

x = Jumlah dari sentimen skor unsur kata dari kalimat

α = Parameter normalisasi yaitu $\alpha = 15$

Normalisasi digambarkan dengan *graph* di bawah ini (Calderon, 2017):



Gambar 1.1. *Graph Normalisasi Metode Analisis Sentimen Vader*

Di sini bahwa ketika x tumbuh lebih besar, ia akan semakin mendekati -1 atau 1. Untuk kasus yang sama, jika ada banyak kata dalam dokumen yang di terapkan analisis sentimen *vader*, akan mendapatkan skor mendekati -1 atau 1. Dengan demikian, analisis sentimen *vader* bekerja paling baik pada dokumen pendek, seperti *tweet*, *instagram* dan kalimat, bukan pada dokumen besar.

Pada penelitian konvensional untuk mengetahui apakah kata-kata tersebut *positif* atau *negatif* (dan secara opsional, sampai tingkat apa), para pengembang pendekatan ini (analisis sentimen) perlu mendapatkan banyak orang untuk menilai mereka secara manual, yang jelas sangat mahal dan memakan waktu. Selain itu, *lexicon* harus memiliki cakupan yang baik dari kata-kata dalam teks yang di minati, selain itu tidak akan sangat akurat. Di sisi lain, ketika ada kecocokan antara *lexicon* dan teks, *vader* merupakan pendekatan yang akurat, dan juga cepat mengembalikan hasil bahkan dalam jumlah besar teks. Untuk itu penggunaan metode *vader* adalah pilihan yang tepat. Pada kasus, pengembang *VADER* digunakan *Mechanical Turk* dari *Amazon* untuk mendapatkan seberapa *rating* mereka, itu adalah merupakan cara yang sangat cepat dan murah untuk mengetahui *rating* mereka (Hutto & Gilbert, 2014). Seperti yang di duga, ketika *VADER* menganalisis sepotong teks, ia memeriksa untuk melihat apakah kata - kata dalam teks ada dalam *lexicon*.

1.6. Sitematika Penulisan

Pada penelitian ini sistematika penulisan dijelaskan pada bagian ini. Berikut sistematika penulisan beserta penjelasannya:

BAB I PENDAHULUAN

Berisi tentang latar belakang penelitian ini, perumusan masalah, tujuan penelitian, ruang lingkup dan batasan masalah, manfaat penelitian, metodologi penelitian, dan penjelasan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Memuat hasil - hasil penelitian sejenis terdahulu yang menginspirasi atau melandasi pelaksanaan penelitian ini, dan juga mengulas landasan teoritik yang berhubungan dengan penelitian yang akan dilakukan.

BAB III RANCANGAN PENELITIAN

Memuat analisis penelitian, meninjau data, mendesain penelitian, serta instrumen penelitian dan juga berisi analisis keperluan alat dan bahan penelitian ini.

BAB IV ANALISIS DAN DISKUSI

Bagian yang memuat berlangsungnya penelitian, pengolahan data, dan menyajikan hasil – hasil yang diperoleh dan cara pencapaiannya serta membahas hasil analisis penelitian ini.

BAB V PENUTUP

Berisi rangkuman hasil penelitian sebagai jawaban rumusan masalah, serta saran yang perlu diperhatikan berdasarkan asumsi penulis.