

LAPORAN
SENTIMENT ANALYSIS WITH TIDY DATA
MENGGUNAKAN BAHASA PEMROGRAMAN “R”



Disusun Oleh :

Niko Fitrianto 16.01.63.0028

UNIVERSITAS STIKUBANK SEMARANG (UNISBANK)

FALKUTAS TEKNOLOGI INFORMASI

TEKNIK INFORMATIKA

2017

BAB I

PENDAHULUAN

1.1 Deskripsi Permasalahan

Penelitian di bidang analisis sentimen dari suatu data merupakan hal yang penting dan dibutuhkan di era big data seperti saat ini. Sentimen positif menyatakan pemberian nilai yang baik, sentimen negatif menyatakan pemberian nilai yang buruk pada konteks dalam bentuk teks, sedangkan sentimen netral menyatakan pemberian nilai yang tidak berpihak pada baik maupun buruk dan biasanya kalimat dengan sentimen ini tidak mengandung opini di dalamnya. Hal ini memacu rasa penasaran suatu korporasi akan sentimen terhadap produk ataupun layanannya biasanya dipacu oleh tingkat persaingan yang semakin tinggi diantara para pelaku pasar. Namun, untuk mengetahui sentimen publik terkadang membutuhkan biaya dan usaha yang tidak mudah. Namun melalui web, perusahaan maupun perorangan dapat mencari jawaban atas keingintahuan tersebut. Opini-opini orang akan berbagai hal dituangkan di web. Yang perlu dilakukan adalah usaha untuk mengumpulkan semua opini tersebut dan mengolahnya menjadi suatu yang dapat menjawab pertanyaan-pertanyaan tadi.

Opini mining atau sentimen analisis adalah riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual. Sekarang ketika suatu organisasi / perusahaan / perorangan ingin memperoleh opini publik mengenai produk, citra dan layanannya, maka mereka tidak perlu melakukan survei konvensional dan fokus group yang mahal biayanya. Deep Learning adalah area baru dalam penelitian Machine Learning, yang telah diperkenalkan dengan tujuan menggerakkan Machine Learning lebih dekat dengan salah satu tujuan aslinya yaitu Artificial Intelligence. Deep Learning adalah tentang belajar beberapa tingkat representasi dan abstraksi yang membantu untuk memahami data seperti gambar, suara, dan teks. Deep Learning mencakup algoritma unsupervised dan supervised learning sehingga dapat memanfaatkan data yang berlabel maupun tidak berlabel.

Analisis sentimen atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Besarnya pengaruh

dan manfaat dari analisis sentimen menyebabkan penelitian dan aplikasi berbasis analisis sentimen berkembang pesat. Bahkan di Amerika terdapat sekitar 20-30 perusahaan yang memfokuskan pada layanan analisis sentimen (Huang, 2009).

Mengeksplorasi secara mendalam apa yang dimaksud dengan format teks yang rapi dan menunjukkan bagaimana format ini dapat digunakan untuk mendekati pertanyaan tentang frekuensi kata. Ini memungkinkan kami menganalisis kata-kata yang paling sering digunakan dalam dokumen dan membandingkan dokumen, namun sekarang mari kita selidiki topik yang berbeda. Ketika pembaca atau manusia mendekati sebuah teks, maka gunakanlah pemahaman kita tentang maksud emosional kata-kata untuk menyimpulkan apakah suatu bagian teks itu positif atau negatif, atau mungkin ditandai oleh emosi lain yang lebih bernuansa seperti kejutan atau jijik. Analisa opini ini bisa saja dilakukan. Misalnya memonitor berita-berita di media masa. Tapi untuk data tweet, cara manual tidak mungkin bisa dilakukan. Ini karena jumlah datanya yang sangat besar dan terus mengalir. Misalnya untuk tweet berbahasa Indonesia saja diperkirakan ada aliran 6 juta tweet per hari (2010). Disinilah peranan text processing, yang secara otomatis dapat memproses data.

Salah satu cara untuk menganalisis sentimen sebuah teks adalah dengan menganggap teks sebagai kombinasi dari kata-kata masing-masing dan isi sentimen keseluruhan teks sebagai jumlah isi sentimen dari masing-masing kata. Ini bukan satu-satunya cara untuk mendekati analisis sentimen, tapi ini adalah pendekatan yang sering digunakan, dan pendekatan yang secara alami memanfaatkan ekosistem alat yang rapi.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas maka rumusan masalahnya adalah bagaimana mengolah data menggunakan *sentiment analysis* dengan *tidy data* beserta implementasinya menggunakan bahasa “R”.

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Deduktif

Opinion Mining / Sentiment Analysis (sebagian besar researcher menganggap dua istilah ini sama/interchangeable) merupakan sebuah cabang penelitian di domain Text Mining yang mulai booming pada awal tahun 2002-an. Riset-nya mulai marak semenjak paper dari B.Pang dan L.Lee [1] keluar. Secara umum, Sentiment analysis ini dibagi menjadi 2 kategori besar :

1. Coarse-grained sentiment analysis
2. Fined-grained sentiment analysis

Coarse-grained sentiment analysis - kita mencoba melakukan proses analysis pada level Dokumen. Singkatnya adalah kita mencoba mengklasifikasikan orientasi sebuah dokumen secara keseluruhan. Orientasi ini ada 3 jenis : Positif, Netral, Negatif. Akan tetapi, ada juga yang menjadikan nilai orientasi ini bersifat kontinu / tidak diskrit.

Fined-grained sentiment analysis merupakan kategori kedua ini yang sedang Naik Daunsekarang. Maksudnya adalah para researcher sebagian besar fokus pada jenis ini. Obyek yang ingin diklasifikasi bukan berada pada level dokumen melainkan sebuah kalimat pada suatu dokumen. Contoh :

1. Saya tidak suka programming. (negatif)
2. Hotel yang baru saja dikunjungi sangat indah sekali. (positif)

Hingga sekarang, hampir sebagian besar penelitian di bidang sentiment analysis hanya ditujukan untuk Bahasa Inggris karena memang Tools/Resources untuk bahasa inggris sangat banyak sekali. Beberapa resources yang sering digunakan untuk sentiment analysis adalah SentiWordNet dan WordNet.

Sentiment analysis terdiri dari 3 subproses besar [2]. Masing-masing subproses ini bisa kita jadikan bahan/topik riset secara terpisah karena masing-masing subproses ini membutuhkan teknik yang tidak mudah :

1. Subjectivity Classification
2. Orientation Detection
3. Opinion Holder and Target Detection

Subjectivity Classification adalah menentukan kalimat yang merupakan opini.

Contoh : “A bike has 2 wheels VS It is a good bike !”

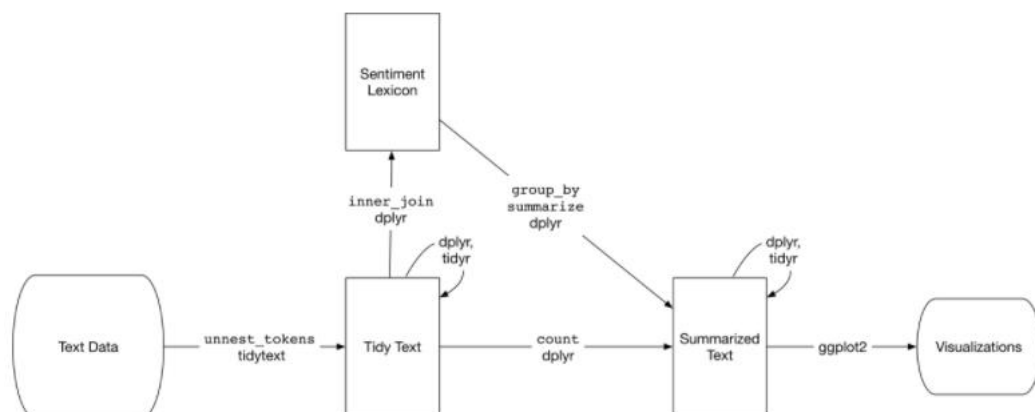
Orientation Detection adalah setelah berhasil diklasifikasi untuk kategori Opini, sekarang kita tentukan apakah dia positif, negatif, netral ?

Contoh : “It is a good bike ! VS ah, It is a bad bike !”

Opinion Holder and Target Detection adalah menentukan bagian yang merupakan Opinion Holder dan bagian yang merupakan Target.

Contoh : “Harry said it is a good bike.”

Ketika pembaca atau manusia mendekati sebuah teks, maka gunakanlah pemahaman kita tentang maksud emosional kata-kata untuk menyimpulkan apakah suatu bagian teks itu positif atau negatif, atau mungkin ditandai oleh emosi lain yang lebih bernuansa seperti kejutan atau jijik. Analisa opini ini bisa saja dilakukan, Misalnya memonitor berita-berita di media masa. Tapi untuk data tweet, cara manual tidak mungkin bisa dilakukan. Ini karena jumlah datanya yang sangat besar dan terus mengalir. Disinilah peranan *text processing*, yang secara otomatis dapat memproses data. Kita bisa menggunakan alat text mining untuk mendekati isi emosional teks secara terprogram, seperti yang ditunjukkan pada Gambar 2.1.



Gambar 2.1. Flowchart Analysis Text

2.2 Kajian Induktif

Tinjauan Pustaka tersebut adalah hasil penelitian terdahulu tentang informasi hasil penelitian yang telah dilakukan sebelumnya dan menghubungkan dengan masalah yang sedang diteliti.

Implementasi text mining pada mesin pencarian twitter untuk mengetahui karakter seseorang menggunakan algoritma Naïve Bayes Classifier. Menurut Ghulam Asrofi Buntoro (2015) dalam jurnal dengan judul Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter, maka dapat ditarik kesimpulan yang menghasilkan sebuah informasi bahwa Analisis Sentimen dapat digunakan untuk mengetahui sentimen masyarakat khususnya netizen Twitter terhadap calon Gubernur DKI Jakarta 2017. Tujuannya membantu masyarakat menentukan sentimen yang terdapat pada tweet opini Bahasa Indonesia yang ada di Twitter. Setelah dilakukan analisis sentimen, terlihat berapa banyak sentimen yang ditujukan kepada calon Gubernur DKI Jakarta 2017. Nilai akurasi tertinggi didapat saat menggunakan metode klasifikasi Naïve Bayes Classifier (NBC) untuk klasifikasi data AHY, dengan nilai rata-rata akurasi mencapai 95%, nilai presisi 95%, nilai recall 95% nilai TP rate 96,8% dan nilai TN rate 84,6%. Dalam penelitian ini juga dapat diketahui metode klasifikasi Naïve Bayes Classifier (NBC) lebih tinggi akurasinya untuk klasifikasi sentimen Tweet Bahasa Indonesia dibandingkan dengan metode klasifikasi Support Vector Machine (SVM).

Pada penelitian lain yang dilakukan oleh Ira Zulfa dan Edi Winarko (2017) dalam jurnal dengan judul Sentimen Analisis Tweet Berbahasa Indonesia dengan *Deep Belief Network*, maka dapat ditarik kesimpulan bahwa Hasil pengujian pada sistem yang dibangun memperlihatkan bahwa metode Deep Belief Network memberikan hasil pengujian klasifikasi lebih baik dengan akurasi sebesar 93.31%, presisi 93%, recall 93%, f1-score 93%, dan support 2378, sedangkan hasil pengujian pada sistem menggunakan metode Naive Bayes memberikan hasil pengujian klasifikasi dengan akurasi sebesar 79.10%, presisi 79%, recall 79%, f1-score 79%, support 2378 dan Support Vector Machine sebesar 92.18%, presisi 92%, recall 92%, f1-score 92% , support 2378. 2. Penggunaan metode DBN dengan menggunakan BagofWord sebagai fitur ekstraksinya terbukti tidak memberikan akurasi lebih baik, ketika dibandingkan dengan metode Naive Bayes dan SVM (Support Vector Machine) dengan menggunakan BagofWord sebagai fitur ekstraksinya pada aplikasi yang dibangun.

BAB III

METODE PENELITIAN

Penelitian ini menggunakan data-data teks pada dataset di dalam *tidy data* yang berkaitan dengan leksikon. Ada berbagai metode dan kamus yang ada untuk mengevaluasi pendapat atau emosi dalam teks. Paket tidytext berisi beberapa leksikon sentimen dalam kumpulan data sentimen. Tujuan akhir penelitian ini adalah dapat menjalankan ke tiga leksikon. Tiga tujuan umum lexicons tersebut adalah :

1. *afinn from Finn Årup Nielsen*.
2. *bing from Bing Liu and collaborators*.
3. *nrc from Saif Mohammad and Peter Turney*.

Ketiganya leksikal didasarkan pada unigrams, yaitu kata tunggal. Leksikon ini mengandung banyak kata dalam bahasa Inggris dan kata-kata tersebut diberi nilai untuk sentimen positif / negatif, dan juga mungkin emosi seperti sukacita, kemarahan, kesedihan, dan sebagainya. Leksikon nrc mengkategorikan kata-kata dengan cara biner ("ya" / "tidak") ke dalam kategori positif, negatif, marah, antisipasi, jijik, ketakutan, sukacita, kesedihan, kejutan, dan kepercayaan. The bing lexicon mengkategorikan kata-kata dalam bentuk biner menjadi kategori positif dan negatif. Leksikon AFINN memberikan kata-kata dengan skor yang berkisar antara -5 dan 5, dengan skor negatif menunjukkan sentimen negatif dan skor positif yang menunjukkan sentimen positif. Semua informasi ini ditabulasikan dalam dataset sentimen, dan tidytext menyediakan fungsi `get_sentiments ()` untuk mendapatkan leksikon sentimen spesifik tanpa kolom yang tidak digunakan dalam leksikon tersebut.

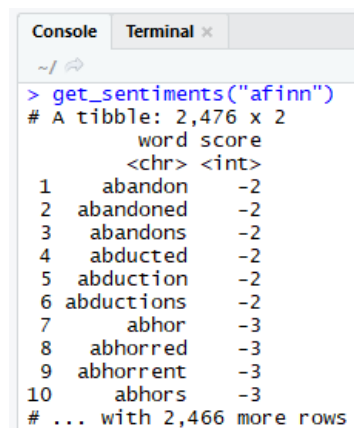
BAB IV

HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

4.1.1. Membaca Data

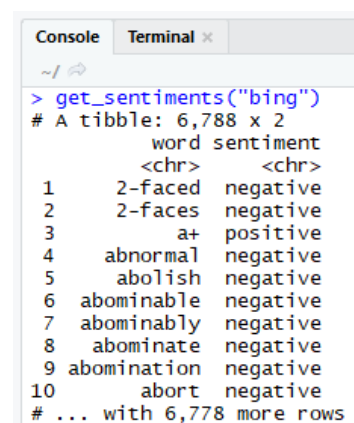
Sebelum melakukan pengolahan data, terlebih dahulu kita abaca data yang akan kita olah nantinya dari ke tiga leksikon tersebut. Dimulai dari melihat data “afinn” seperti yang ditampilkan pada Gambar 4.1.



```
Console Terminal x
~/
> get_sentiments("afinn")
# A tibble: 2,476 x 2
  word score
  <chr> <int>
1 abandon -2
2 abandoned -2
3 abandons -2
4 abducted -2
5 abduction -2
6 abductions -2
7 abhor -3
8 abhorred -3
9 abhorrent -3
10 abhors -3
# ... with 2,466 more rows
```

Gambar 4.1. Data “afinn”

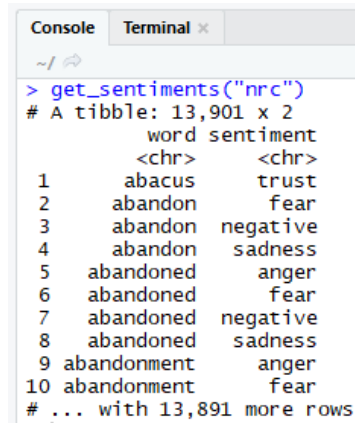
Leksikon AFINN memberikan kata-kata dengan skor yang berkisar antara -5 dan 5, dengan skor negatif menunjukkan sentimen negatif dan skor positif yang menunjukkan sentimen positif. Kemudian dilanjutkan dengan melihat data “bing” seperti yang ditampilkan pada Gambar 4.2.



```
Console Terminal x
~/
> get_sentiments("bing")
# A tibble: 6,788 x 2
  word sentiment
  <chr> <chr>
1 2-faced negative
2 2-faces negative
3 a+ positive
4 abnormal negative
5 abolish negative
6 abominable negative
7 abominably negative
8 abominate negative
9 abomination negative
10 abort negative
# ... with 6,778 more rows
```

Gambar 4.2. Data “bing”

Leksikon bing mengkategorikan kata-kata dalam bentuk biner menjadi kategori positif dan negatif. Kemudian dilanjutkan dengan melihat data “nrc” seperti yang ditampilkan pada Gambar 4.3.



```
> get_sentiments("nrc")
# A tibble: 13,901 x 2
  word      sentiment
  <chr>    <chr>
1  abacus      trust
2  abandon     fear
3  abandon    negative
4  abandon    sadness
5  abandoned   anger
6  abandoned   fear
7  abandoned   negative
8  abandoned   sadness
9  abandonment anger
10 abandonment fear
# ... with 13,891 more rows
```

Gambar 4.3. Data “nrc”

Leksikon nrc mengkategorikan kata-kata dengan cara biner ("ya" / "tidak") ke dalam kategori positif, negatif, marah, antisipasi, jijik, ketakutan, sukacita, kesedihan, kejutan, dan kepercayaan.

Semua informasi ini ditabulasikan dalam dataset sentimen, dan tidytext menyediakan fungsi `get_sentiments()` untuk mendapatkan leksikon sentimen spesifik tanpa kolom yang tidak digunakan dalam leksikon tersebut. Bagaimana leksikon sentimen ini disatukan dan divalidasi? Mereka dibangun melalui crowdsourcing (menggunakan, misalnya, Amazon Mechanical Turk) atau oleh tenaga kerja dari salah satu penulis, dan divalidasi menggunakan beberapa kombinasi crowdsourcing lagi, ulasan restoran atau film, atau data Twitter. Dengan informasi ini, kita mungkin ragu untuk menerapkan leksikon sentimen ini ke gaya teks yang secara dramatis berbeda dari apa yang telah mereka divalidasi, seperti fiksi narasi dari 200 tahun yang lalu. Meskipun benar bahwa menggunakan leksikon sentimen ini dengan, misalnya, novel Jane Austen mungkin memberi kita hasil yang kurang akurat daripada dengan tweet yang dikirim oleh penulis kontemporer, kita masih dapat mengukur konten sentimen untuk kata-kata yang dibagi di seluruh teks leksikon dan teks.

4.2. Pengolahan Data

4.2.1. Sentiment Analysis with Inner Join

Lihat kata-kata dengan skor sukacita dari kamus NRC. Apa kata-kata sukacita yang paling umum di Emma? Pertama, kita perlu mengambil teks dari novel dan mengubah teks menjadi format rapi menggunakan `unnest_tokens()`. Mari kita juga membuat beberapa kolom lain untuk melacak garis dan bagian dari buku yang masing-masing

berasal dari kata, kami menggunakan `group_by` dan `bermutasi` untuk membangun kolom tersebut dengan kode program sebagai berikut :

```
library(tidytext)
get_sentiments("afinn")
get_sentiments("bing")
library(janeaustenr)
library(dplyr)
library(stringr)

tidy_books <- austen_books() %>%
group_by(book) %>%
mutate(linenumber = row_number(),
chapter = cumsum(str_detect(text, regex("^chapter [\\divxlc]", ignore_case =
TRUE)))) %>%
ungroup() %>%
unnest_tokens(word, text)

nrcjoy <- get_sentiments("nrc") %>%
filter(sentiment == "joy")

tidy_books %>%
filter(book == "Emma") %>%
inner_join(nrcjoy) %>%
count(word, sort = TRUE)
```

```
Joining, by = "word"
# A tibble: 303 x 2
  word      n
  <chr> <int>
1 good   359
2 young  192
3 friend 166
4 hope   143
5 happy  125
6 love   117
7 deal   92
8 found   92
9 present 89
10 kind   82
# ... with 293 more rows
```

Gambar 4.4. Result of Join by Word

```
library(tidyr)
```

```
janeaustensentiment <- tidy_books %>%
```

```
inner_join(get_sentiments("bing")) %>%
```

```
count(book, index = linenumber %/% 80, sentiment) %>%
```

```
spread(sentiment, n, fill = 0) %>%
```

```
mutate(sentiment = positive - negative)
```

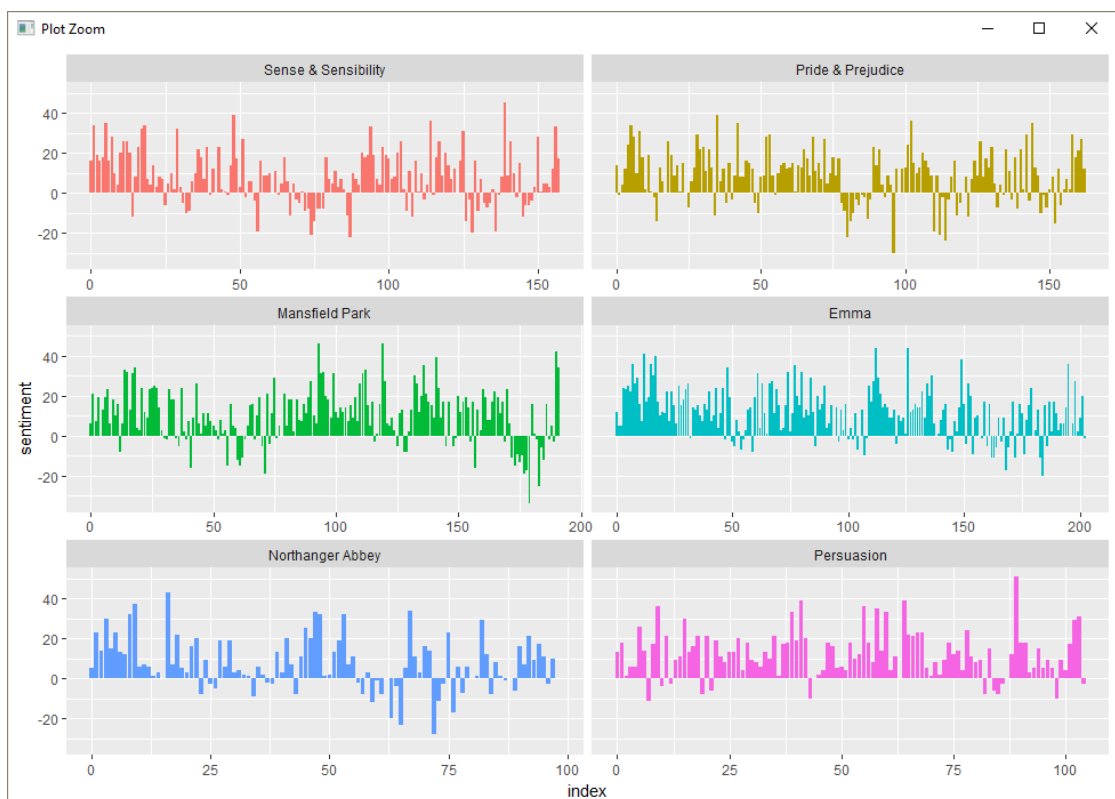
```
library(ggplot2)
```

```
ggplot(janeaustensentiment, aes(index, sentiment, fill = book)) +
```

```
geom_col(show.legend = FALSE) +
```

```
facet_wrap(~book, ncol = 2, scales = "free_x")
```

Hasil :



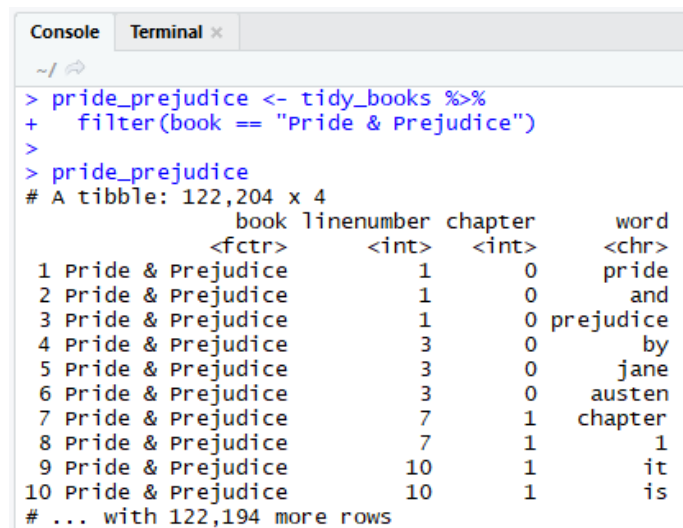
Gambar 4.5. *Sentiment through the narratives of Jane Austen's novels*

Kita dapat melihat pada Gambar 4.5 bagaimana plot setiap novel berubah menjadi sentimen positif atau negatif lebih jauh dari lintasan cerita.

4.2.2. Comparing the Three Sentiment Dictionaries

Dengan beberapa pilihan untuk leksikon sentimen, Anda mungkin menginginkan beberapa informasi lebih lanjut tentang mana yang sesuai untuk tujuan Anda. Mari kita gunakan ketiga leksikon sentimen tersebut dan periksa bagaimana sentimen berubah melintasi busur naratif *Pride and Prejudice*. Pertama, ayo gunakan filter () untuk memilih hanya kata-kata dari satu novel yang kami minati.

```
pride_prejudice <- tidy_books %>%  
filter(book == "Pride & Prejudice")  
pride_prejudice
```



```
> pride_prejudice <- tidy_books %>%  
+   filter(book == "Pride & Prejudice")  
>  
> pride_prejudice  
# A tibble: 122,204 x 4  
   book linenumber chapter word  
   <fctr>      <int>   <int> <chr>  
1 Pride & Prejudice      1     0 pride  
2 Pride & Prejudice      1     0 and  
3 Pride & Prejudice      1     0 prejudice  
4 Pride & Prejudice      3     0 by  
5 Pride & Prejudice      3     0 jane  
6 Pride & Prejudice      3     0 austen  
7 Pride & Prejudice      7     1 chapter  
8 Pride & Prejudice      7     1 1  
9 Pride & Prejudice     10     1 it  
10 Pride & Prejudice     10     1 is  
# ... with 122,194 more rows
```

Gambar 4.6. Hasil filter buku dengan nama *Pride & Prejudice*

```
afinn <- pride_prejudice %>%  
inner_join(get_sentiments("afinn")) %>%  
group_by(index = linenumber %/% 80) %>%  
summarise(sentiment = sum(score)) %>%  
mutate(method = "AFINN")  
  
bing_and_nrc <- bind_rows(pride_prejudice %>%  
inner_join(get_sentiments("bing")) %>%  
mutate(method = "Bing et al."),  
pride_prejudice %>%
```

```

inner_join(get_sentiments("nrc") %>%
  filter(sentiment %in% c("positive", "negative"))) %>%
mutate(method = "NRC") %>%
count(method, index = linenumber %% 80, sentiment) %>%
spread(sentiment, n, fill = 0) %>%
mutate(sentiment = positive - negative)

bind_rows(afinn,
bing_and_nrc) %>%
ggplot(aes(index, sentiment, fill = method)) +
geom_col(show.legend = FALSE) +
facet_wrap(~method, ncol = 1, scales = "free_y")

```

Hasil :



Gambar 4.7. Comparing three sentiment lexicons using *Pride and Prejudice*

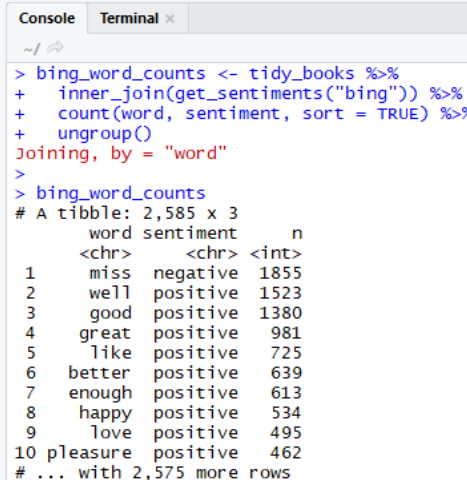
Tiga leksikon yang berbeda untuk menghitung sentimen memberikan hasil yang berbeda dalam arti absolut namun memiliki lintasan relatif serupa melalui novel ini.

Kita melihat penurunan dan puncak yang sama dalam sentimen di tempat yang sama dalam novel, namun nilai absolutnya berbeda secara signifikan. Leksikon AFINN memberikan nilai absolut terbesar, dengan nilai positif tinggi. Leksikon dari Bing et al. memiliki nilai absolut yang lebih rendah dan tampaknya memberi label blok yang lebih besar dari teks positif atau negatif bersebelahan. Hasil NRC bergeser lebih tinggi dibandingkan dengan dua lainnya, memberi label teks lebih positif, namun mendeteksi perubahan relatif serupa dalam teks. Kami menemukan perbedaan yang sama antara metode saat melihat novel lainnya; Sentimen NRC tinggi, sentimen AFINN memiliki varian yang lebih banyak, Bing et al. Sentimen tampaknya menemukan peregangan panjang dari teks serupa, namun ketiganya sepakat secara kasar mengenai keseluruhan tren dalam sentimen melalui busur naratif.

4.2.3. Most Common Positive and Negative Words

Salah satu keuntungan memiliki bingkai data dengan sentimen dan kata adalah bahwa kita dapat menganalisis jumlah kata yang berkontribusi pada setiap sentimen. Dengan menerapkan hitungan () di sini dengan argumen dari kedua kata dan sentimen, kita mengetahui berapa banyak setiap kata berkontribusi pada setiap sentimen, dengan menggunakan kode program berikut :

```
bing_word_counts <- tidy_books %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE) %>%
  ungroup()
bing_word_counts
```



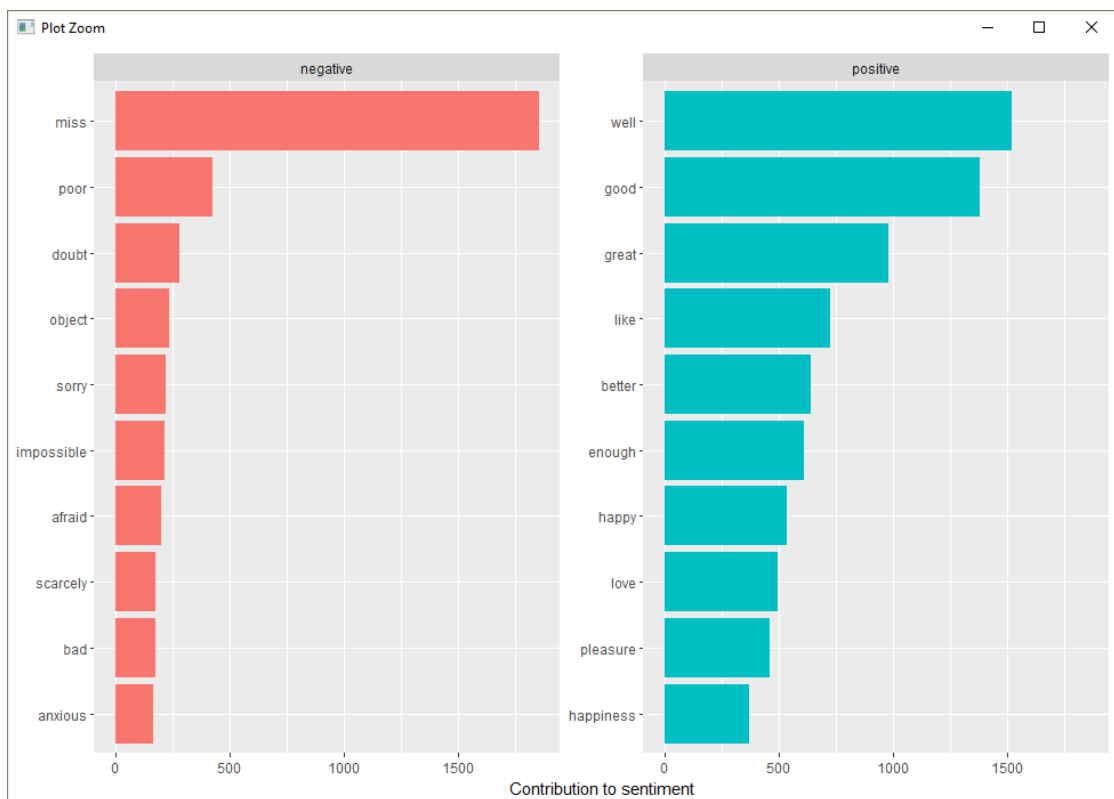
```
Console Terminal x
~/
> bing_word_counts <- tidy_books %>%
+   inner_join(get_sentiments("bing")) %>%
+   count(word, sentiment, sort = TRUE) %>%
+   ungroup()
Joining, by = "word"
>
> bing_word_counts
# A tibble: 2,585 x 3
   word sentiment      n
  <chr>      <chr> <int>
1 miss    negative 1855
2 well    positive 1523
3 good    positive 1380
4 great   positive  981
5 like    positive  725
6 better   positive  639
7 enough   positive  613
8 happy    positive  534
9 love     positive  495
10 pleasure positive  462
# ... with 2,575 more rows
```

Gambar 4.8. Result of Join by Word

```

bing_word_counts %>%
group_by(sentiment) %>%
top_n(10) %>%
ungroup() %>%
mutate(word = reorder(word, n)) %>%
ggplot(aes(word, n, fill = sentiment)) +
geom_col(show.legend = FALSE) +
facet_wrap(~sentiment, scales = "free_y") +
labs(y = "Contribution to sentiment", x = NULL) + coord_flip()

```



Gambar 4.8. *Contributor to Sentiment*

Gambar 4.8 memungkinkan kita melihat anomali dalam analisis sentimen; kata "miss" dikodekan sebagai negatif tapi digunakan sebagai judul untuk wanita muda yang belum menikah dalam karya Jane Austen. Jika sesuai untuk tujuan kita, kita dapat dengan mudah menambahkan "rindu" ke daftar kata kunci stop-kata dengan menggunakan `bind_rows()`. Kita bisa menerapkannya dengan strategi seperti ini.

Ukuran teks kata pada Gambar 4.11, sebanding dengan frekuensinya dalam sentimennya. Kita dapat menggunakan visualisasi ini untuk melihat kata-kata positif dan negatif yang paling penting, namun ukuran kata-katanya tidak sebanding dengan sentimen.

BAB V

KESIMPULAN DAN REKOMENDASI

Opini mining atau sentimen analisis adalah riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual. Sekarang ketika suatu organisasi / perusahaan / perorangan ingin memperoleh opini publik mengenai produk, citra dan layanannya, maka mereka tidak perlu melakukan survei konvensional dan fokus group yang mahal biayanya. Analisis sentimen atau opinion mining merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Ada berbagai metode dan kamus yang ada untuk mengevaluasi pendapat atau emosi dalam teks. Paket `tidytext` berisi beberapa leksikon sentimen dalam kumpulan data sentimen. Leksikon “nrc” mengkategorikan kata-kata dengan cara biner (“ya” / “tidak”) ke dalam kategori positif, negatif, marah, antisipasi, jijik, ketakutan, sukacita, kesedihan, kejutan, dan kepercayaan. Leksikon “bing” mengkategorikan kata-kata dalam bentuk biner menjadi kategori positif dan negatif. Leksikon “afinn” memberikan kata-kata dengan skor yang berkisar antara -5 dan 5, dengan skor negatif menunjukkan sentimen negatif dan skor positif yang menunjukkan sentimen positif. Semua informasi ini ditabulasikan dalam dataset sentimen, dan `tidytext` menyediakan fungsi `get_sentiments ()` untuk mendapatkan leksikon sentimen spesifik tanpa kolom yang tidak digunakan dalam leksikon tersebut.

DAFTAR PUSTAKA

- Ira Zulfa, Edi Winarko. 2017. Sentimen Analisis Tweet Berbahasa Indonesia dengan Deep Belief Network. IJCCS, Vol.11, No.2, July 2017, pp. 187-198.
- Ghulam Asrofi Buntoro. 2016. Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter. Integer Journal, Vol 1, No 1, Maret 2016: 32-41.
- Ismail Sunni, Dwi Hendratmo Widyantoro. 2012. Analisis Sentimen dan Ekstraksi Topik Penentu Sentimen pada Opini Terhadap Tokoh Publik. Jurnal Sarjana Institut Teknologi Bandung Bidang Teknik Elektro dan Informatika. Volume 1, Number 2, Juli 2012.
- Zhao, Yanchang 2011. R and Data Mining: Sentiment Analysis with Tidy Data. Julia Sigle and David Robinson. Sumber : <https://www.tidytextmining.com/sentiment.html#wordclouds> (27 Desember 2017).