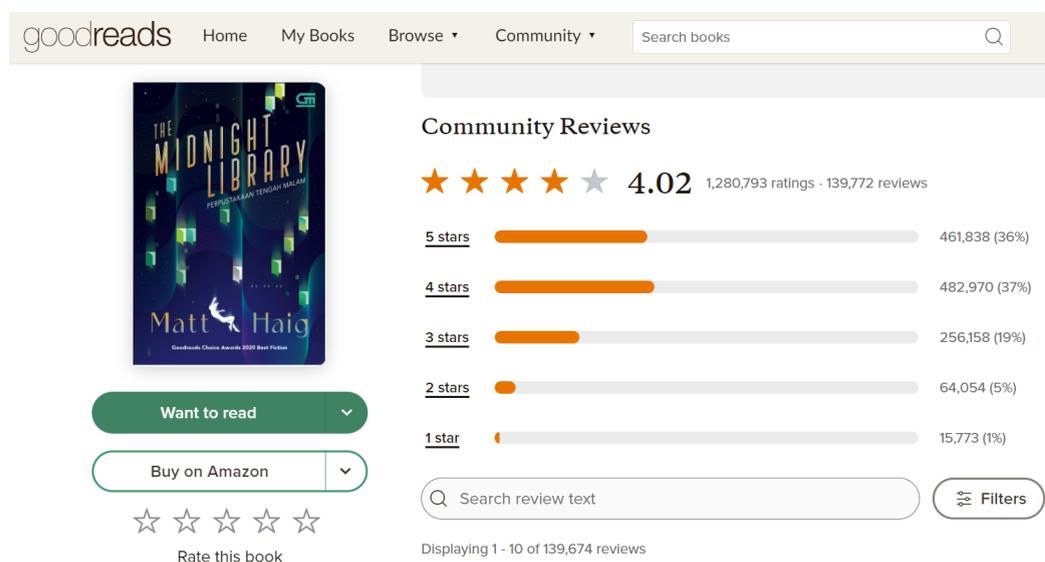


BAB IV

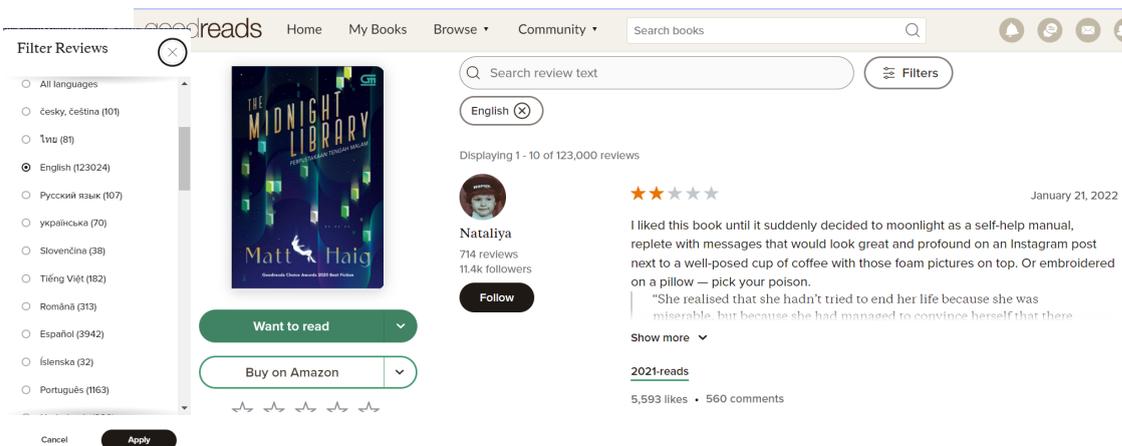
HASIL PENELITIAN DAN PEMBAHASAN

A. Gambaran Umum Review Pembaca



Gambar 4.1 Gambaran Umum Review

Secara keseluruhan, per-tanggal 09 Februari 2023 novel *The Midnight Library* karya Matt Haig memiliki rating 4.02 dari 1.280.793 pemberi *ratings* dan 139.772 pemberi review novel. Gambar diatas menunjukkan bahwa sebagian besar pembaca memberi *rating* 4 dan 5 pada novel ini. Dengan kata lain, dapat disebutkan bahwa mayoritas pembaca menyukainya. Rating tersebut diberikan oleh komunitas pembaca di situs *Goodreads.com* sejak novel ini diterbitkan pertama kali dalam bahasa Inggris dan dicetak ulang dalam bentuk terjemahan ke beberapa bahasa lain untuk didistribusikan secara global.



Gambar 4.2 Jumlah Review (*English*)

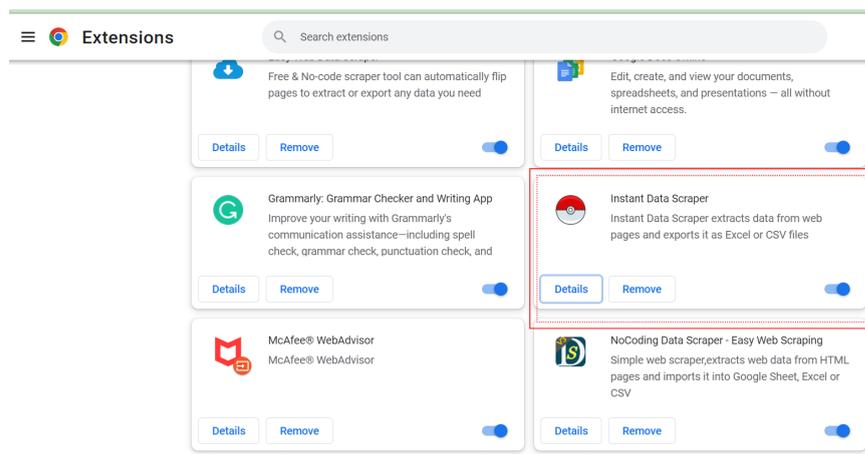
Mengingat *Goodreads.com* adalah komunitas pembaca buku yang anggotanya tersebar luas secara global. Ulasan buku terdiri dari bahasa yang sangat bervariasi dari seluruh dunia. Namun, pengunjung situs dapat memanfaatkan fitur *filter* untuk mengetahui review pembaca pada bahasa tertentu. Pada Gambar 4.2 dapat terlihat bahwa jumlah review berbahasa Inggris ada sebanyak 123.000. Banyaknya jumlah data ini akan memudahkan proses pengumpulan data agar nantinya dapat diperoleh jumlah yang seimbang nantinya antara positif, negatif, dan netral untuk pengolahan data.

B. Klasifikasi Review Pembaca untuk Analisis Sentimen

1. Pengumpulan Data

Proses pengumpulan data pada penelitian ini dilakukan dengan mengambil data review dari pengguna situs pada website *Goodreads.com*. Proses pengambilan data dilakukan dengan menggunakan teknik *web scraping* (mengambil data dari sebuah halaman web menjadi format data terstruktur yang dapat disimpan ke dalam database untuk keperluan analisis lebih lanjut) menggunakan Google Chrome *extension* Instant Data Scraper versi 1.0.8.

Terdapat beberapa langkah pengerjaan dalam melakukan *scraping data*. Langkah pertama adalah menginstal aplikasi dan mengaktifkan ekstensi pada Google Chrome seperti yang ditunjukkan pada Gambar 4.3 di bawah ini.



Gambar 4.3 Ekstensi Data Miner

Ekstensi Data Miner adalah perangkat tambahan atau *add-on* untuk browser web seperti Google Chrome dan Mozilla Firefox yang memungkinkan pengguna untuk mengambil data dari halaman web dan menganalisisnya dengan alat analisis data. Ekstensi ini sangat berguna bagi peneliti untuk mengumpulkan informasi dari internet dan menggunakannya untuk penelitian. Dengan Ekstensi Data Miner, pengguna dapat dengan mudah mengekstrak data dari tabel atau grafik pada halaman web dan menyimpannya dalam format yang dapat diolah oleh alat analisis data seperti Excel atau Google Sheets.

Pada penelitian ini terdapat sumber data yang digunakan berupa review yang diberikan oleh pembaca novel *The Midnight Library* pada situs *Goodreads*. Berikut ini alamat domain dari website *Goodreads* untuk diambil data reviewnya:

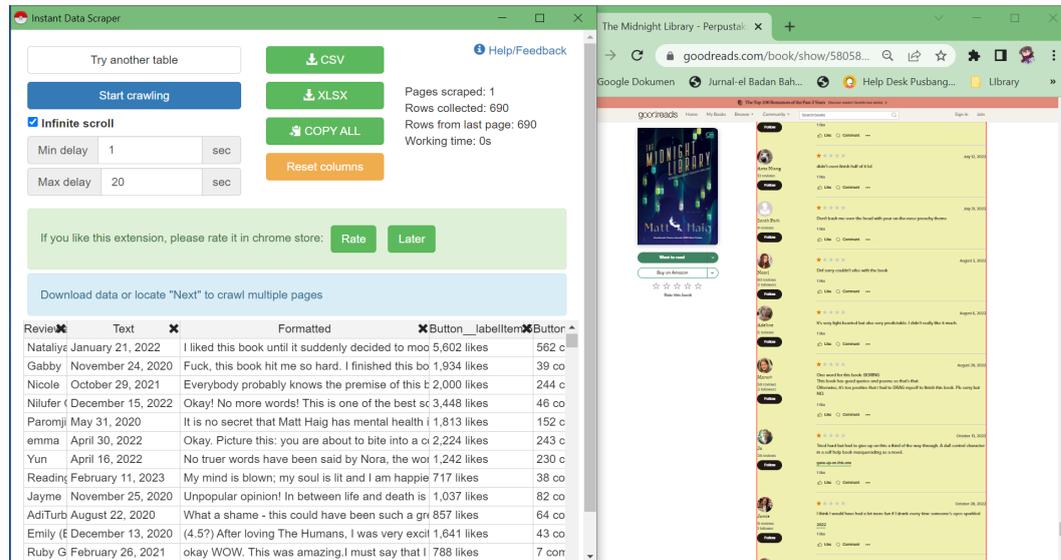
<https://www.Goodreads.com/book/show/58058734-the-midnight-library---perpustakaan-tengah-malam>.

Langkah selanjutnya adalah membuka situs web yang alamatnya diberikan di atas, seperti yang terlihat pada Gambar 4.4.



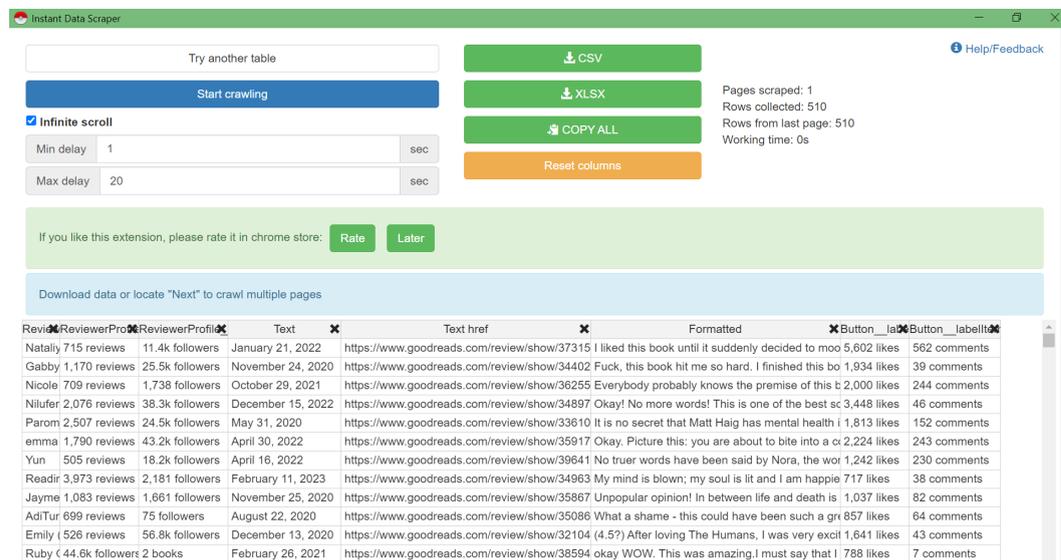
Gambar 4.4 Website *Goodreads*

Tahap ini melakukan pengambilan data sesuai dengan data yang dibutuhkan. Pertama lakukan pemilihan atribut mana yang ingin diambil datanya. Peneliti mengarahkan Instant Data Scraper untuk mengambil data review pengguna. Selanjutnya klik, “*Start crawling*” untuk mulai mengekstrak data pada website.



Gambar 4.5 Mengekstrak Data

Pada tahap ini, harus dipastikan komponen penting berupa review pembaca sudah berhasil diekstrak semua sebelum melanjutkan ke langkah berikutnya.



Gambar 4.6 Hasil Crawling pada Instant Data Scraper

Lalu, periksa kolom apa saja yang akan digunakan dalam penelitian seperti yang dilakukan pada Gambar 4.6. Pada tahap ini, ada banyak data tidak terpakai

yang juga ikut terambil. Misalnya adalah tautan *follow*, *like*, dan *comment*. Maka harus dipastikan memilah data yang tidak sesuai dan tidak menyertakannya saat proses *download*. Dengan menekan tombol download, file dengan ekstensi CSV atau Excel yang telah disimpan setelah proses scraping data selesai dilakukan. File tersebut dapat dibuka dengan aplikasi seperti Microsoft Excel, Google Sheets, atau aplikasi pengolah data lainnya. Setelah file dibuka, kita dapat melihat hasil scraping data yang telah dilakukan seperti ditunjukkan oleh Gambar 4.7 berikut.

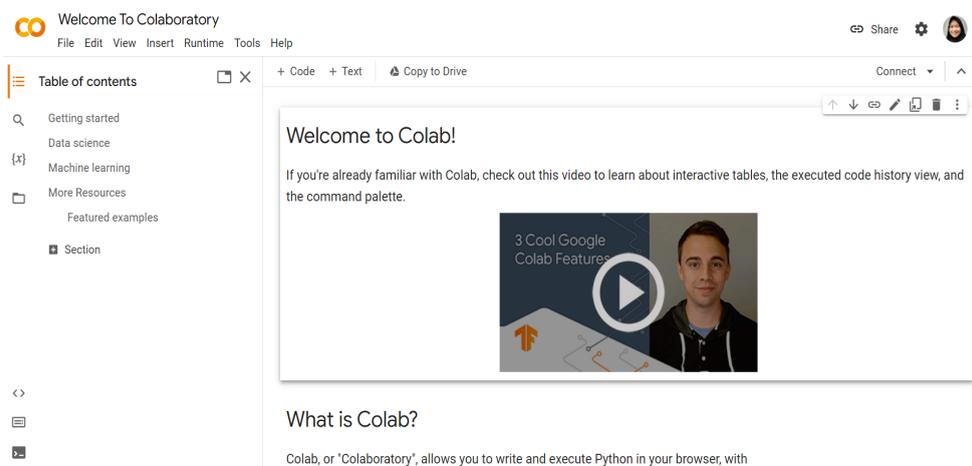
| | A | B | C | D |
|----|-------------------|-------------------|----------------------------------------------------------------------------------------------------------------------------------------|------|
| 1 | ReviewerProfile | Text | Formatted | Star |
| 2 | Nataliya | January 21, 2021 | I liked this book until it suddenly decided to moonlight as a self-help manual, replete with messages that would look great and | 1 |
| 3 | Gabby | November 24, 2021 | Fuck, this book hit me so hard. I finished this book a few days ago and I haven't been able to stop thinking about it. It has one | 1 |
| 4 | Nicole | October 29, 2021 | Everybody probably knows the premise of this book by now (I mean this book is): a library with an infinite number of books. T | 1 |
| 5 | Nilufer Ozmekik | December 15, 2021 | Okay! No more words! This is one of the best sci-fi dances with fantasy which carries additional philosophic vibes novel of the | 1 |
| 6 | Paromjit | May 31, 2020 | It is no secret that Matt Haig has mental health issues, dogged by the darkness of depression that has taken its toll on his life. | 1 |
| 7 | emma | April 30, 2022 | Okay. Picture this: you are about to bite into a cookie. A big, warm cookie. Kinda crispy on the outside, chewy on the inside, y | 1 |
| 8 | Yun | April 16, 2022 | No truer words have been said by Nora, the world's largest wet blanket. Woe is her, folks. Nora ends up at a place between li | 1 |
| 9 | Reading_Tamisi | February 11, 2021 | My mind is blown; my soul is lit and I am happier."She realised, in that moment, that she was capable of a lot more than she h | 1 |
| 10 | Jayme | November 25, 2021 | Unpopular opinion! In between life and death is a library. The Midnight Library. With books curated just for you..You will start w | 1 |
| 11 | AdiTurbo | August 22, 2020 | What a shame - this could have been such a great novel if the author had the capacity to write about the issues of regrets anc | 1 |
| 12 | Emily (Books wit | December 13, 2021 | 4.5? After loving The Humans, I was very excited for this new release and I wasn't let down!This books was heartbreaking ai | 1 |
| 13 | Ruby Granger | February 26, 2021 | okay WOW. This was amazing.I must say that I was kind of skeptical? going into this because the idea is SO good that I didn't | 1 |
| 14 | Hailey (Hailey in | February 1, 2021 | I think this is a really deeply personal book, and it's definitely one that resonated with me. It is kind of part self help book and p | 1 |
| 15 | Cindy | August 19, 2021 | Corny like a Hallmark movie and probably the least subtle book I've ever read.Update: for those butthurt about my sentence-lc | 1 |
| 16 | Emily B | March 29, 2021 | This was cute and the concept was great but unfortunately it really lacked some depth for me! | 1 |

Gambar 4.7 Hasil *Web Scraping*

Melalui proses *scraping* yang telah dilakukan, peneliti berhasil mengumpulkan beragam review pembaca yang bernilai positif maupun negatif sebanyak 3.000 review untuk novel *The Midnight Library*. Data tersebut dibaca menggunakan Google Spreadsheet untuk memudahkan proses *review dataset*, yakni mencari tahu adanya baris yang kosong dan pembersihan data awal.

2. Gambaran Umum Alat Penelitian

Peneliti menggunakan data yang didapatkan pada proses sebelumnya untuk proses pengolahan data review pembaca novel *The Midnight Library* karya Matt Haig. Berikut adalah alat yang digunakan peneliti.



Gambar 4.8 Tampilan Awal Google Colaboratory

Google Colaboratory (<https://colab.research.google.com/>), merupakan salah satu produk dari Google Internal Research berupa *compiler* dan editor bahasa pemrograman Python secara *online* yang dapat digunakan secara gratis. Situs <https://research.google/> sendiri aktif melakukan publikasi artikel penelitian mengenai teknologi *computer science*.

Peneliti memilih menggunakan Google Colab karena dapat digunakan menyimpan, menulis, serta membagikan kode Python tanpa perlu melakukan proses instalasi dan *setup* lainnya. Terutama juga karena saat ini Python telah menjadi sangat populer dan merupakan bahasa yang sangat baik untuk belajar

tentang komputasi.⁴⁵ Google Colab serupa dengan Jupyter Notebook gratis berbentuk *cloud* yang dijalankan menggunakan *browser*, seperti Mozilla Firefox dan Google Chrome. Google Colab cukup mudah digunakan jika ingin mempelajari dan mengasah pengetahuan mengenai bahasa pemrograman Python.

Kode yang telah disusun juga dapat disimpan pada Google Drive karena Google Colaboratory ini berbasis *cloud* sehingga tidak memakai banyak memori memori internal komputer. Selain penyimpanan, proses saat menjalankan program juga melalui *cloud* sehingga pengguna tidak dituntut memiliki spesifikasi komputer yang tinggi untuk menjalankan program ini.

Keberadaan *library* Python sebagai modul berisi kumpulan kode yang dapat digunakan berulang kali dalam program yang berbeda membuat pemrograman Python menjadi lebih sederhana dan nyaman karena tidak perlu menulis kode yang sama berulang kali untuk program yang berbeda untuk pembelajaran mesin, *data science*, dan visualisasi data. Saat ini, ada lebih dari 137.000 *library* Python yang mengeliminasi kebutuhan untuk menulis kode dari awal. *Library* Python sangat memudahkan dalam proses penelitian karena peneliti dapat melakukan proses pengolahan data dari awal sampai akhir hanya menggunakan satu bahasa pemrograman untuk pengolahan data pada Google Colab.

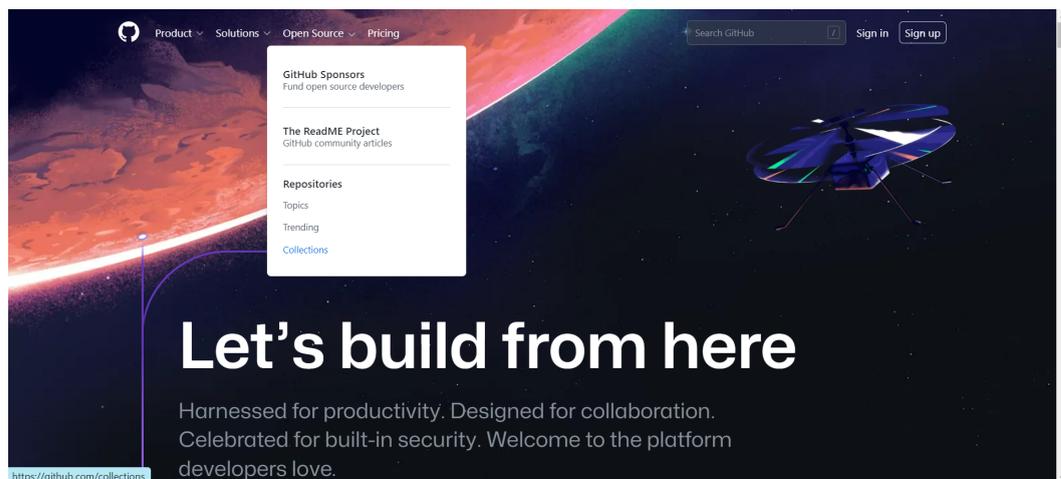
3. *Pre Processing Data*

Data yang diambil dari web scraping pada tahap sebelumnya masih belum dapat digunakan karena masih memiliki bentuk yang tidak terstruktur. Oleh karena itu, diperlukan proses pra-pemrosesan data untuk mempersiapkan data agar

⁴⁵ Cioffi-Revilla, Claudio. "Introduction to Computational Social Science: Principles and Applications." Second Edition. *London and Heidelberg: Springer* (2017).

siap digunakan. Tahap *preprocessing* ini dilakukan menggunakan Google Colaboratory untuk menjalankan kode pemrograman Python.

Sebelumnya, *code* tersebut peneliti kutip dari situs GitHub. Situs ini menyediakan *code* yang bisa digunakan secara *open source* (kode yang dirancang agar dapat diakses publik—siapa pun dapat melihat, memodifikasi, dan mendistribusikan kode sesuai keinginan) karena GitHub menawarkan layanan *hosting repository* berbasis *cloud*. Dengan demikian, pengguna bisa *posting* hasil *project/repo/source code* agar *user* lain dapat memberikan *like*, saran, dan bahkan bisa memodifikasinya.



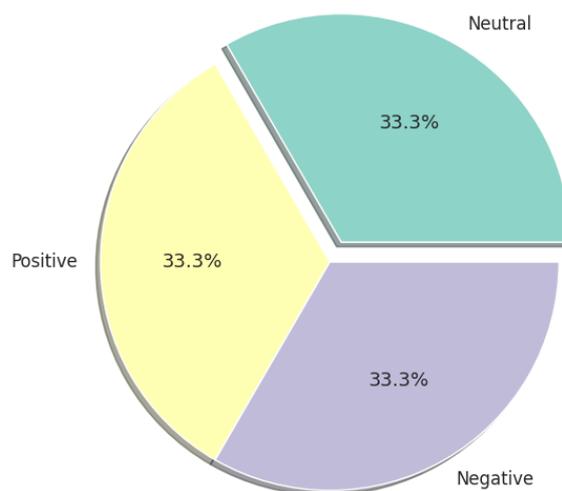
Gambar 4.9 Tampilan Awal GitHub

Gambar 4.9 menunjukkan tampilan awal situs GitHub sebelum peneliti mengakses *repository* kode Python untuk analisis sentimen. Seperti halnya sebuah situs *repository* untuk jurnal, pengguna bisa mengakses publikasi terdahulu dan mengutip beberapa bagian sesuai *project* yang sedang dikerjakan. Keberadaan *source code* analisis sentimen yang dipublikasikan secara *open source* sangat

memudahkan peneliti dalam melakukan penelitian ini. *Code* tersebut disediakan oleh pengguna dengan *username* Arpan-dev⁴⁶ yang merupakan seorang *data science enthusiast*. Selain itu, peneliti juga menggunakan referensi tambahan dari video demonstrasi yang diunggah mahasiswa program studi Informatika Universitas Bengkulu mengenai analisis sentimen menggunakan SVM.⁴⁷

Setelah mendapatkan *source code* yang sesuai untuk penelitian ini. Peneliti menggunakannya pada Google Colaboratory untuk memastikan jumlah data yang seimbang antara sentimen negatif, netral, maupun positif. Sebagaimana dapat terlihat pada grafik lingkaran di atas, kita dapat mengetahui bahwa terdapat 33.3% sentimen negatif, 33.3% sentimen positif, dan 33.3% sentimen netral dari 3.000 data yang akan diolah. Distribusi data penelitian dapat dilihat pada grafik berikut dalam bentuk diagram lingkaran.

Distribution of sentiments



Gambar 4.10 Distribusi Data

⁴⁶ Arpan-dev. *Sentiment Analysis: sentiment_analysis_project_clg.ipynb*, 2022. Diakses dari <https://github.com/Arpan-dev/Sentiment-Analysis/>

⁴⁷ Afrizah, Yogi, & Afrianda, Deri. *Analisis Sentimen Menggunakan Metode Support Vector Machine*. Video Youtube, 2020. Diakses dari <https://www.youtube.com/watch?v=3DJ3mCIKNWA>

Adapun tahapan *preprocessing* yang akan dilakukan adalah sebagai berikut.

a. *Case Folding*

Tahap *case folding*, juga dikenal sebagai perubahan huruf besar ke huruf kecil, digunakan untuk membuat teks review lebih terstruktur dengan menyeragamkan semua kata menjadi huruf kecil dan menghilangkan tanda baca. Hal ini bertujuan agar kata-kata yang sebenarnya memiliki arti yang sama namun ditulis dengan huruf besar dan kecil tidak dianggap berbeda. Meskipun tidak ada aturan pasti mengenai tahapan dalam *text preprocessing*, melakukan semua tahapan dapat memastikan hasil yang lebih baik dan konsisten. Semakin banyak tahapan yang dilakukan, semakin dalam lapisan data yang dapat dikupas.

b. *Tokenizing*, proses pemisahan teks menjadi potongan kata (*token*).

c. *Stopword Removal*

Tahapan ini bertujuan menghilangkan kata-kata yang tidak memiliki pengaruh signifikan pada kalimat dan tidak mengurangi informasi atau makna dari kalimat. Proses penghapusan *stopword* menghemat waktu dan ruang memori yang digunakan, dan tidak akan mengurangi efektivitas dari informasi yang diambil.

Tabel 4.1 Transformasi Teks

| | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <ul style="list-style-type: none"> - Lower case conversion - Word Tokenization - Alphanumeric Checking - Stopword Removal - Perform Stemming | <pre>def transform_text(text): text=text.lower() text=nlk.word_tokenize(text) L=[] for i in text: if i.isalnum(): L.append(i) text=L[:] #copying list L.clear()</pre> |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

| | |
|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | <pre> for i in text: if i not in pos_sw and i not in string.punctuation: L.append(i) text=L[:] L.clear() for i in text: L.append(ps.stem(i)) return " ".join(L) </pre> |
|--|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Code pada Tabel 4.1 ditulis dalam bahasa pemrograman Python untuk memberikan perintah transformasi teks review pada Google Colaboratory sebagai proses *text preprocessing* atau pembersihan data. Perintah yang diberikan mencakup mengubah semua kata pada review menjadi huruf kecil (*lower case conversion*), mengubah kalimat menjadi potongan kata agar memudahkan pembobotan kata (*word tokenization*), memeriksa kombinasi karakter alfabet dan numerik (*alphanumeric checking*), menghapus kata yang tidak akan memberi makna signifikan untuk analisis sentimen, seperti kata ganti dan kata hubung (*stopword removal*), kemudian menghapus imbuhan pada bagian depan maupun belakang dari kata (*stemming*).

Tabel 4.2 Hasil Transformasi Teks

| url_rmvd_text | trans_text |
|---------------------------------------------------|---------------------------------------------------|
| Honestly, the more I think about this, the mor... | honestli think realiz 3 star enjoy would still... |
| A reflective read that does a fairly good job ... | reflect read fairli good job flesh idea though.. |

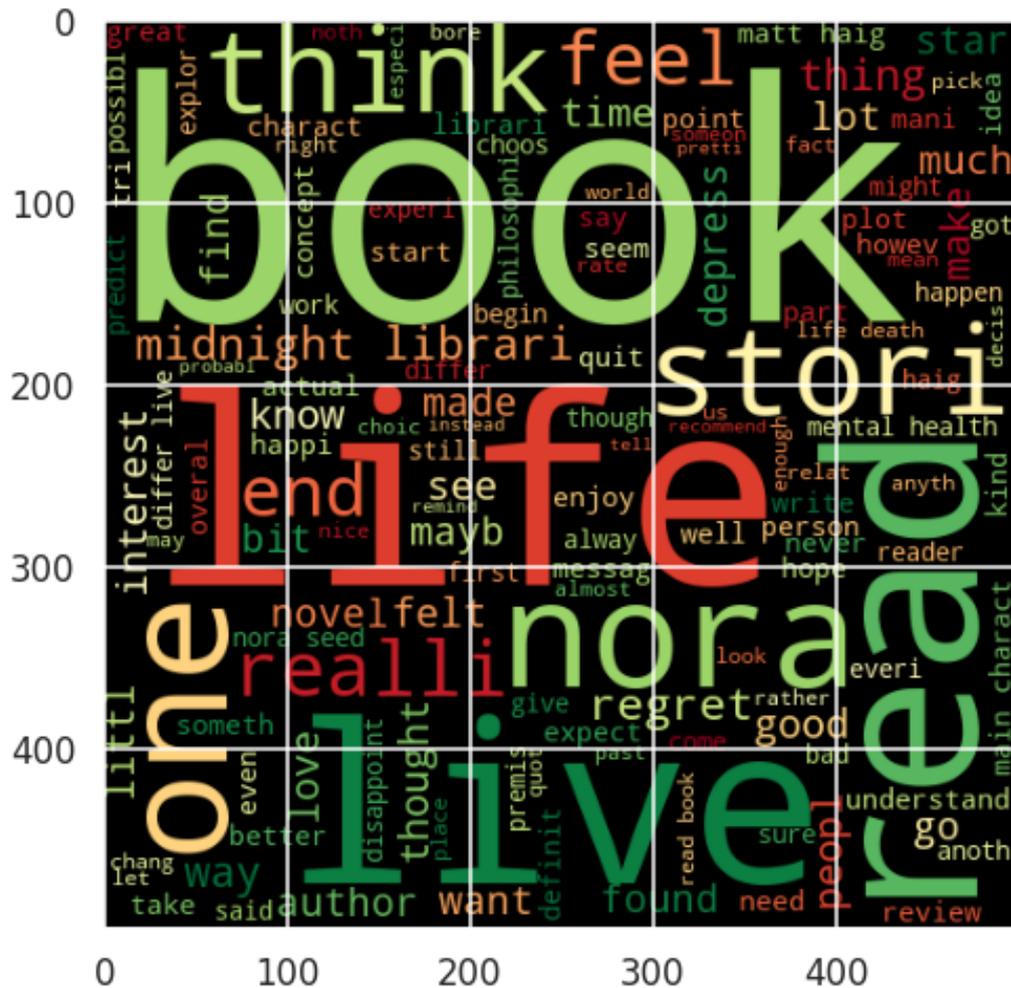
| | |
|------------------------------------------------------|-----------------------------------------------------|
| “you don’t have to understand life, you just h... | understand life live bit beach read ml moment .. |
| Another book that had potential to become a fa... | anoth book potenti becom favorit mine unfortun |

Berdasarkan tabel di atas, dapat terlihat bahwa ada beberapa kata yang terhapus, seperti kata ganti dan kata sambung pada bagian kiri tabel (teks berwarna merah). Lalu bagian sebelah kanan sebagai hasil transformasi teks hanya menampilkan kata-kata dasar saja. Namun demikian, dapat terlihat beberapa kata yang terpotong bagian ujungnya karena proses *stemming* (menghilangkan imbuhan pada kata). Pada proses ini, peneliti tidak menggunakan proses *lemmatization* (penyesuaian kata sesuai kata dasar yang terdapat pada kamus).

4. Visualisasi *Word Cloud*

Word cloud adalah representasi visual dari data teks untuk menggambarkan kata kunci di situs web, atau memvisualisasikan teks bentuk bebas. Kata-kata yang direpresentasikan berbentuk tunggal dengan dilakukan penyesuaian ukuran atau warna font sesuai banyaknya kemunculan kata pada dataset review. Hal ini dilakukan untuk menyoroti hanya beberapa kata-kata yang nilainya penting. Grafik ini peneliti dapatkan dengan menjalankan kode Python di Google Colaboratory melalui *library* Matplotlib untuk melakukan visualisasi grafis.

a. Sentimen Netral



Gambar 4.11 Visualisasi Sentimen Netral

Gambar 4.11 di atas menunjukkan visualisasi sentimen netral yang diambil dari hasil *text mining* 1.000 review pembaca berbintang 3 (tiga) dari situs *Goodreads.com*. Ulasan tersebut yang berbentuk kalimat telah dipecah menjadi kata-kata berbentuk tunggal. Kata-kata seperti, “book”, “life”, “live”, “nora”, “one”, dan “read” memiliki ukuran huruf yang lebih besar dibandingkan kata-kata lainnya. Hal ini berarti kata-kata tersebut sering muncul pada review pembaca yang berisi sentimen netral.

dengan adanya kata, “regret”, “depress”, “suicide”, dan “disappoint”. Keinginan “suicide” diawali berbagai kekecewaan Nora sehingga memutuskan mengakhiri hidup. Pembaca khawatir hal ini dapat menjadi *trigger* bagi orang lain yang sedang mengalami situasi serupa, yakni memiliki banyak penyesalan-penyesalan.

Awalnya pembaca memiliki ekspektasi novel ini akan menceritakan petualangan menyenangkan para tokohnya di perpustakaan saat tengah malam. Namun bahasan novel ini lebih mengarah pada topik *self-help*, yakni tindakan atau proses memperbaiki diri sendiri atau mengatasi masalah seseorang tanpa bantuan orang lain. Oleh karena itu, ada banyak buku *self-help* yang beredar berisi kata-kata motivasi bagi pembaca agar tetap kuat menjalani hidup. Pada novel ini ditunjukkan dengan kemampuan Nora bangkit dari depresinya. Kekecewaan pembaca karena ketidaksesuaian ekspektasi dengan isi novel menyebabkan rasa kecewa terhadap novel. Hal ini direpresentasikan dengan *keyword*, “disappoint” pada *word cloud*.

Secara keseluruhan, sentimen negatif pada novel ini disebabkan tidak sesuai harapan atau ekspektasi pembaca saat membeli dengan konten teks cerita yang mereka baca pada novel. Berkaitan dengan konten atau isi pada novel, alur cerita tentang penyesalan, depresi, dan keinginan bunuh diri yang terdapat pada novel ini juga termasuk kata kunci yang sering muncul dari review pembaca terhadap novel *The Midnight Library*.

netral dan negatif. Sedangkan kata bersentimen positif pada visualisasi ini dapat ditemukan dengan adanya kata, “good”, “love”, “happy”, “help”, “perfect”, “beauty”, “appreciate”, dan “mental health”. Kata kunci tersebut selaras dengan isi novel *The Midnight Library* bahwa Nora Seed bertemu dengan Ms. Elm selaku pustakawan di perpustakaan tengah malam. Ms. Elm sering memberinya nasihat-nasihat bijak mengenai cara menghadapi penyesalan karena pilihan yang telah diambil di masa lalu.

Sentimen positif berkaitan dengan pembaca yang memahami alasan tokoh utama digambarkan sebagai orang yang depresi karena memiliki terlalu banyak penyesalan. Pembaca yakin bahwa hal ini tidak terlepas dari unsur ekstrinsik novel berupa subjektivitas pengarang, yakni Matt Haig sendiri yang merupakan seorang penyintas isu kesehatan mental. Beberapa karya terkait dari Matt adalah buku berjudul *Reason to Stay Alive* yang diangkat dari kisah yang pernah dialami penulis. Buku ini menceritakan tentang perjalanan Matt dalam menghadapi depresinya ketika masih muda, bagaimana perasaannya dan cara dia bangkit.

Selanjutnya, buku *The Comfort Book* yang berisi perenungan dari Matt Haig setelah mengalami depresi, selain itu terdapat kisah inspiratif dari beberapa tokoh. Ada juga topik seputar kehidupan bahkan kematian untuk membuka cakrawala pembaca. Lalu buku *How to Stop Time* yang membuat pembaca merenungkan kembali kehidupan yang dijalani. Kutipan yang terkenal dari buku ini adalah, “Selama bertahun-tahun ini aku berhasil meyakinkan diri bahwa kepedihan kenangan terasa lebih berat dan lama daripada saat-saat bahagia”.

5. Klasifikasi

Sebelum proses klasifikasi sentimen, dilakukan pembagian data latih dan data uji. Data latih akan dijadikan acuan oleh komputer untuk mempelajari pola kata-kata yang terdapat pada kalimat review novel untuk menentukan jenis sentimen yang terkandung di dalamnya. Penelitian ini melakukan pembagian data dengan 80% data latih dan 20% data uji. Tujuan diperbanyak data latih adalah memberi kesempatan pada komputer untuk mempelajari lebih banyak pola kata-kata yang terkandung pada review novel.

Tabel 4.3 Pembagian Data Latih dan Data Uji

| | |
|---------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Data Train and Data Test Split</i> | <pre>from sklearn.model_selection import train_test_split X_train,X_test,y_train,y_test=train_test_split(X,y,test_size =0.2,random_state=42)</pre> |
|---------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------|

Tabel 4.3 di atas adalah perintah kode yang dijalankan oleh peneliti untuk memisahkan data latih (*data train*) dan data uji (*data test*). Hal ini dilakukan dengan memanggil *library* sklearn pada Python. *Library* ini sangat berguna untuk proses pembelajaran mesin (*machine learning*) karena dapat menjadi alat yang efisien untuk membuat model klasifikasi data.

Setelah membagi data menjadi data uji dan data testing sesuai dengan skenario yang telah ditentukan, langkah selanjutnya adalah melakukan klasifikasi menggunakan metode *Support Vector Machine*. Pada penelitian ini juga dilakukan klasifikasi beberapa algoritma lainnya sebagai pembandingan.

Tabel 4.4 Membuat Model Klasifikasi

| | |
|-------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Creating Model for MultinomialNB, BernoulliNB, Random forest, Logistic regression, Decision tree, and Support Vector Machine</i> | <pre> from sklearn.naive_bayes import MultinomialNB,BernoulliNB from sklearn.ensemble import RandomForestClassifier from sklearn.linear_model import LogisticRegression from sklearn.tree import DecisionTreeClassifier from sklearn.svm import LinearSVC from sklearn.metrics import accuracy_score,precision_score,multil abel_confusion_matrix,recall_score </pre> |
|-------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

Tabel 4.4 di atas adalah rangkaian kode yang digunakan pada penelitian ini untuk menjalankan perintah pembuatan model klasifikasi. Peneliti kembali menggunakan *library* sklearn untuk memanggil algoritma SVM dan beberapa algoritma lainnya. Algoritma yang dijadikan pembanding juga lazim dipergunakan peneliti lain pada penelitian terdahulu untuk klasifikasi sentimen dari sebuah data berbentuk teks.

Tabel 4.5 Hasil Klasifikasi

| Algoritma | Akurasi | Persentase |
|---------------------------------|----------------|-------------------|
| <i>Logistic Regression</i> | 0.775000 | 77,5% |
| <i>Support Vector Machine</i> | 0.773333 | 77,3% |
| <i>Multinomial Naive Bayes</i> | 0.755000 | 75,5% |
| <i>Random Forest Classifier</i> | 0.746667 | 74,7% |
| <i>Decision Tree Classifier</i> | 0.681667 | 68,1% |
| <i>Bernoulli Naive Bayes</i> | 0.653333 | 65,3% |

Berdasarkan Tabel 4.5, terlihat bahwa algoritma *Logistic Regression* dan *Support Vector Machine* memiliki nilai akurasi yang tinggi dengan persentase akurasi masing-masing adalah 77,5% dan 77,3% untuk 80% data latih dan 20% data uji dari total 3.000 review novel. Urutan tersebut disusul oleh algoritma *Multinomial Naive Bayes* dan *Random Forest Classifier* dengan tingkat akurasi 75,5%. Lalu *Decision Tree Classifier* memiliki akurasi 68,1%. Terakhir adalah algoritma *Bernoulli Naive Bayes* dengan persentase akurasi sebanyak 65,3%.

6. Evaluasi

Tahap evaluasi dilakukan dengan membandingkan persentase akurasi klasifikasi sentimen review menggunakan *Support Vector Machine* (SVM) dengan beberapa algoritma lainnya. Berdasarkan Tabel 4.5, dapat dilihat bahwa SVM memang memiliki akurasi yang baik dengan persentase 77,3%. Hal ini sesuai dengan hasil penelitian terdahulu yang peneliti temukan pada tugas akhir maupun terbitan artikel jurnal yang merekomendasikan penggunaan SVM sebagai algoritma yang bagus untuk klasifikasi.

Beberapa di antaranya adalah penelitian dari Salwa Kamilia Hasna tahun 2021 mengenai, “Analisis Sentimen Data Ulasan Menggunakan Algoritma *Support Vector Machine* (Studi Kasus: Aplikasi Iflix)” yang memperoleh hasil proses klasifikasi menggunakan metode SVM dengan akurasi tertinggi pada skenario data latih 70% dan data uji 30% sebesar 93,45%.⁴⁸ Kemudian penelitian Ragil Dimas Himawan dan Eliyani pada tahun 2021 yang berjudul, “Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi

⁴⁸ Hasna, Salwa Kamilia. "Analisis Sentimen Data Ulasan Menggunakan Algoritma Support Vector Machine (Studi Kasus: Aplikasi Iflix)." Skripsi, Universitas Islam Indonesia, 2021.

DKI Jakarta di Masa Pandemi". Penelitian tersebut menyatakan terdapat perbedaan tingkat akurasi pada penggunaan tiga metode klasifikasi yaitu Linear SVM, Multinomial Naive Bayes, dan Random Forest Classifier dalam mengklasifikasikan data tweet masyarakat terhadap Pemerintah DKI Jakarta. Dari ketiga metode tersebut, Linear SVM memiliki akurasi paling tinggi (77,58%).⁴⁹

Adanya nilai persentase akurasi hasil klasifikasi sentimen ini memungkinkan peneliti menghitung berapa jumlah review yang berhasil terklasifikasi secara benar dari data uji. Sebagaimana disebutkan pada kajian teori, rumus *confusion matrix* yang dapat digunakan adalah sebagai berikut.

$$Akurasi = \frac{TN + TP}{TN + TP + FN + FP}$$

Peneliti dapat menggunakan rumus akurasi pada *confusion matrix* ini untuk mencari tahu jumlah review yang dapat diklasifikasikan secara benar dengan menggunakan hasil persentase akurasi yang telah didapat setelah sebelumnya saat menjalankan kode pemrograman Python untuk melakukan pemodelan SVM pada data uji. Oleh karena pada penelitian ini menggunakan 80% data latih dan 20% data uji pada 3.000 review pada dataset, maka jumlah data uji yang digunakan dapat dihitung seperti berikut.

⁴⁹ Himawan, Ragil Dimas, and Eliyani Eliyani. "Perbandingan Akurasi Analisis Sentimen Tweet terhadap Pemerintah Provinsi DKI Jakarta di Masa Pandemi." *JEPIN (Jurnal Edukasi dan Penelitian Informatika)* 7, no. 1 (2021): 58-63.

$$\frac{100\%}{20\%} = \frac{3.000}{\text{Jumlah data uji}}$$

$$(\text{Jumlah data uji}) 100 = 3.000 \times 20$$

$$(\text{Jumlah data uji}) 100 = 60.000$$

$$\text{Jumlah data uji} = \frac{60.000}{100}$$

$$\text{Jumlah data uji} = 600$$

Berdasarkan perhitungan tersebut, dapat diketahui bahwa terdapat 600 review yang digunakan sebagai data uji saat melakukan klasifikasi sentimen. Jumlah tersebut meliputi review yang berhasil klasifikasi benar ataupun salah (TN + TP + FN + FP) saat dilakukan pemodelan oleh SVM. Berikut ini hasil penghitungan yang dilakukan untuk mencari tahu berapa jumlah review yang berhasil terklasifikasi secara benar (TN + TP).

$$\text{Akurasi} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}}$$

$$77,3\% = \frac{\text{TN} + \text{TP}}{\text{Jumlah data uji}}$$

$$\frac{77,3}{100} = \frac{\text{TN} + \text{TP}}{600}$$

$$(\text{TN} + \text{TP}) 100 = 77,3 \times 600$$

$$(\text{TN} + \text{TP}) 100 = 46.380$$

$$\text{TN} + \text{TP} = \frac{46.500}{100}$$

$$\text{TN} + \text{TP} = 463,8$$

Berdasarkan penghitungan di atas, dapat diketahui bahwa terdapat 463,8 atau dapat dibulatkan menjadi 464 review yang berhasil diklasifikasikan secara benar untuk pemodelan SVM pada penelitian ini. Selain itu, peneliti juga melakukan uji

coba dengan memasukkan sembarang review dari dataset. Lalu dijalankan sebuah perintah yang meminta komputer menentukan apakah teks tersebut memiliki sentimen negatif, netral, atau positif. Perintah tersebut dilakukan dengan menjalankan kode di bawah ini.

Tabel 4.6 *Blind Data*

```

tweet = [input("Enter a Review: ")]

tweet = pd.DataFrame(tweet,columns = ["Reviews"])
tweet

url_rmvd_text = url_rmv(tweet["Reviews"])
url_rmvd_text

tweet["url_rmvd_text"]= url_rmvd_text
#tweet

tweet["trans_text"]=tweet["url_rmvd_text"].apply(transform_text)
#tweet

cv = pickle.load(open('final_cv.pickle','rb'))

X = cv.transform(tweet["trans_text"])

cv = pickle.load(open('final_model.pickle','rb'))
pred=loaded_model.predict(X)[0]

if pred == 0:
    print("Negative Sentiment")
elif pred == 1:
    print("Neutral Sentiment")
else:
    print("Positive Sentiment")

```

Melalui kode ini, peneliti dapat memasukkan salah satu review secara acak yang bersumber dari dataset. Perintah tersebut dilakukan saat muncul tulisan, “Enter a Review: ”. Lalu berdasarkan hasil transformasi teks, komputer diminta mengidentifikasi apakah review tersebut bernilai negatif, netral, atau positif.

C. *Text Mining* sebagai Kajian di Bidang Data dan Informasi

Text mining adalah proses penggalian data yang dilakukan sistem komputer pada teks untuk mendapatkan informasi yang berguna dari sekumpulan data atau dokumen. Tujuan akhir dari *text mining* adalah untuk menghasilkan data yang dapat diolah menjadi informasi yang bermanfaat. Data yang dihasilkan dari *text mining* biasanya berupa teks yang tidak terstruktur dan tidak memiliki format yang jelas, sehingga perlu dilakukan tahap *preprocessing* agar dapat dikelompokkan atau dikategorikan.⁵⁰ Oleh karena itu, *text mining* bermaksud mendapatkan pengetahuan baru yang tersembunyi dari sekumpulan teks.

Tahapan terakhir proses *text mining* adalah *text analytic*. Data teks yang sudah bersih dan diidentifikasi berdasarkan *term/kata* kunci menjadi ciri teks tersebut telah diolah menggunakan algoritma untuk kebutuhan analisis. *Text analytics* berupa analisis sentimen adalah pendekatan untuk mengestimasi klasifikasi teks dalam berbagai macam sentimen (positif, negatif, dan netral). *Term* tersebut peneliti dapatkan setelah pengolahan data di Google Colaboratory dan direpresentasikan dalam bentuk Word Cloud yang telah ditampilkan sebelumnya pada pembahasan mengenai klasifikasi review.

Pengklasifikasian review pembaca novel *The Midnight Library* menggunakan metode *Support Vector Machine* (SVM) telah dilakukan dan menghasilkan persentase akurasi sebanyak 77,3% dengan 80% data latih dan 20% data uji pada 3.000 review pembaca berbahasa Inggris. Lebih lanjut, sentimen juga ditunjukkan pada visualisasi *term/kata* kunci.

⁵⁰ Hans Christian, Pius. "Analisis Sentimen Deteksi Ulasan Palsu Terhadap Toko Online Menggunakan Algoritma Naive Bayes, Support Vector Machine, dan Logistic Regression." Skripsi, Universitas Multimedia Nusantara, 2022.

Berdasarkan analisis sentimen pembaca novel yang sudah dilakukan, peneliti merekomendasikan perpustakaan umum dan perguruan tinggi memiliki novel *The Midnight Library* sebagai salah satu koleksi karena tidak hanya dapat menjadi bacaan hiburan atau rekreasi dengan membaca fiksi, tetapi juga mengandung nilai pembelajaran dan amanat yang bisa diambil oleh pembaca. Sambutan baik dari pembaca direpresentasikan pada visualisasi *word cloud* sentimen positif berupa, “good”, “love”, “happy”, “help”, “perfect”, “beauty”, “appreciate”, dan “mental health”.

Pada perpustakaan perguruan tinggi, novel ini dapat menjadi bahan penelitian untuk mahasiswa dari berbagai program studi seperti psikologi, bahasa, hingga ilmu perpustakaan dan informasi. Namun, terkhusus pada perpustakaan umum yang memiliki karakteristik pembaca beragam dari berbagai rentang usia, pustakawan dapat membatasi akses bagi pemustaka anak-anak yang ingin membaca dan meminjamnya karena topik bahasan yang diangkat oleh buku ini termasuk pada kategori dewasa, yakni berkaitan dengan penyesalan-penyosalan hidup tokoh utama, depresi, hingga adanya percobaan bunuh diri. Versi terjemahan Indonesia dari Penerbit Gramedia sendiri memberi informasi bahwa buku ini diperuntukkan pada usia 17+.

Selanjutnya dalam kaitan hasil visualisasi data dengan ilmu perpustakaan sebagai kajian di bidang data dan informasi dapat ditunjukkan menggunakan diagram *fishbone* (diagram tulang ikan) atau yang juga sering disebut *Cause and Effect diagram* adalah diagram yang menyerupai tulang ikan untuk menunjukkan sebab akibat dari suatu permasalahan. *Fishbone* diagram digunakan ketika kita

ingin mengidentifikasi kemungkinan penyebab apa saja dari suatu masalah atau akibat yang ditimbulkannya. Dengan demikian, *fishbone* akan mengkategorikan berbagai sebab potensial dari satu masalah atau pokok persoalan dengan cara yang mudah dimengerti, rapi, dan terstruktur.

Fishbone akan menggambarkan hubungan antara masalah dengan semua faktor penyebab yang mempengaruhi masalah tersebut. Jenis diagram ini juga disebut diagram "Ishikawa" karena ditemukan oleh Kaoru Ishikawa, seorang ahli kimia yang berasal dari Jepang. Dengan mengelompokkan faktor penyebab masalah, diagram ini memungkinkan pengklasifikasian berbagai sumber variasi ke dalam kategori dan membantu mempersempit akar penyebab masalah. Bagian kepala ikan merepresentasikan masalah/akibat, bagian tulang merepresentasikan penyebab utama, dan bagian rusuk tulang merepresentasikan sub-penyebab.⁵¹

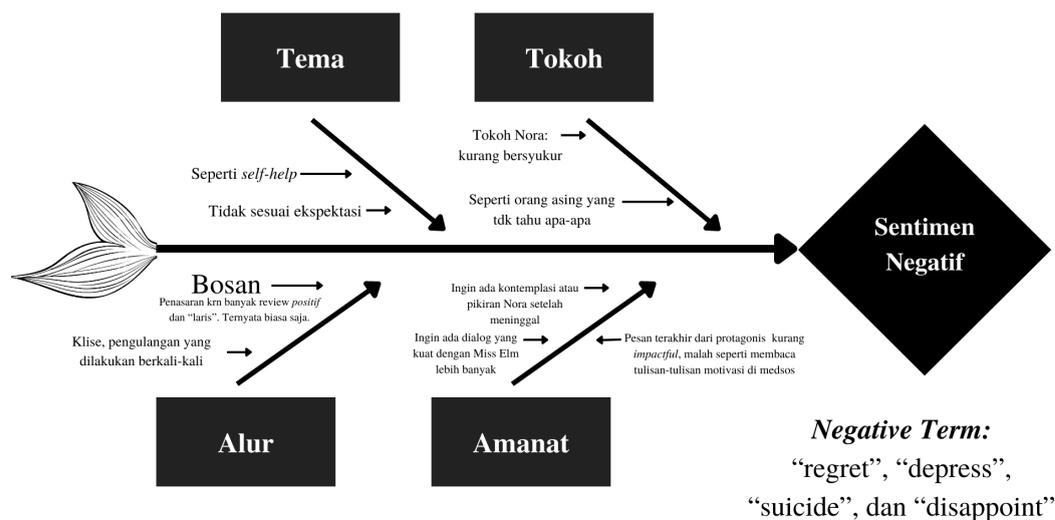
Diagram ini membawa manfaat untuk perbaikan proses serta pemecahan masalah berkat kelengkapan visualnya, yakni dengan membantu mempresentasikan secara visual pernyataan masalah atau keberadaan suatu akibat, penyebab utama, serta sub-penyebab pendukungnya. Selain itu, diagram ini juga mendukung ide-ide *brainstorming* dengan menjadi alat grafis yang memberikan pandangan tentang keadaan masalah yang ada.⁵²

Fishbone pada penelitian ini dibuat sebagai visualisasi sentimen positif dan negatif untuk menganalisa penyebab review baik maupun buruk novel *The Midnight Library* berdasarkan data yang ada. Berkaitan dengan hal ini, sebagai

⁵¹ Anirudha, Anastasya. "Popular Articles: Mengenal Fishbone Diagram". Halaman web, 2022. Diakses dari <https://sis.binus.ac.id/2022/08/26/mengenal-fishbone-diagram/>

⁵² _____. "Bahan Ajar Universitas Dian Nuswantoro: Diagram Fishbone". Diakses dari https://repository.dinus.ac.id/docs/ajar/Diagram_Fisbone.pdf.

acuan utama dari penyebab sentimen, peneliti memilih menggunakan beberapa unsur-unsur intrinsik novel yang sempat disebutkan pada kajian teori sebagai tulang ikan yang akan dijabarkan sub penyebabnya pada bagian rusuk tulang.



Gambar 4.14 Diagram *Fishbone* Sentimen Negatif

Diagram *fishbone* untuk sentimen negatif di atas menyoroti pendapat pembaca mengenai tema, tokoh, alur, dan amanat yang terkandung dalam novel *The Midnight Library* karya Matt Haig.

1. Tema

Kebanyakan *reviewer* di *Goodreads* mengategorikan novel ini sebagai fantasi dan realisme magis. Realisme magis merupakan aliran sastra yang menggunakan pendekatan realis/kejadian sehari-hari dengan gabungan elemen magis di dalamnya ditandai dengan hadirnya hal-hal magis di dunia nyata yang akan menghancurkan konsep ruang dan waktu itu sendiri. Realisme magis adalah aliran sastra yang menggunakan pendekatan realis/kejadian sehari-hari dengan

gabungan elemen magis di dalamnya. Sedangkan cerita fantasi berarti dunia yang tidak nyata, aneh, dan ada penjelasan tentang kenapa keanehan itu terjadi.

Nora Seed sebagai tokoh utama, dapat mengubah pilihan dia ambil di masa lalu dan menjalani berbagai kehidupan yang bisa dia dapatkan. Nora dapat mengulang kembali pilihan hidupnya selama tengah malam belum berakhir, dengan harapan menemukan pilihan yang benar-benar diinginkannya. Dengan demikian, *reviewer* yang bersentimen negatif mengenai tema merasa cerita novel ini tidak sesuai dengan tema yang dia harapkan.

Namun pembaca berpendapat bahwa alih-alih novel ini menceritakan realisme magis atau fantasi yang mereka bayangkan, penulis justru membuat novel ini tampak seperti buku *self-help* mengenai isu kesehatan mental.

2. Tokoh

Reviewer yang memberi sentimen negatif mengenai tokoh merasa bahwa Nora Seed kurang bersyukur dalam menjalani kehidupannya. Apalagi setelah mengulang banyak kehidupan, pada akhirnya dia hanya kembali pada kehidupan lama yang sebelumnya ingin dia tinggalkan. Selain itu, karakter pada tokoh Nora juga dianggap kurang kuat dan tidak memberi kontribusi mendalam dengan hanya bersikap seperti orang yang tidak tahu apa-apa selama berulang kali mencoba berbagai kehidupan. Selanjutnya, penulis juga kurang mengeksplor peranan tokoh-tokoh lain seperti Ms. Elm (pustakawan) dan Hugo Lefevre yang juga menjadi penjelajah kehidupan seperti Nora (*slider* yang lain).

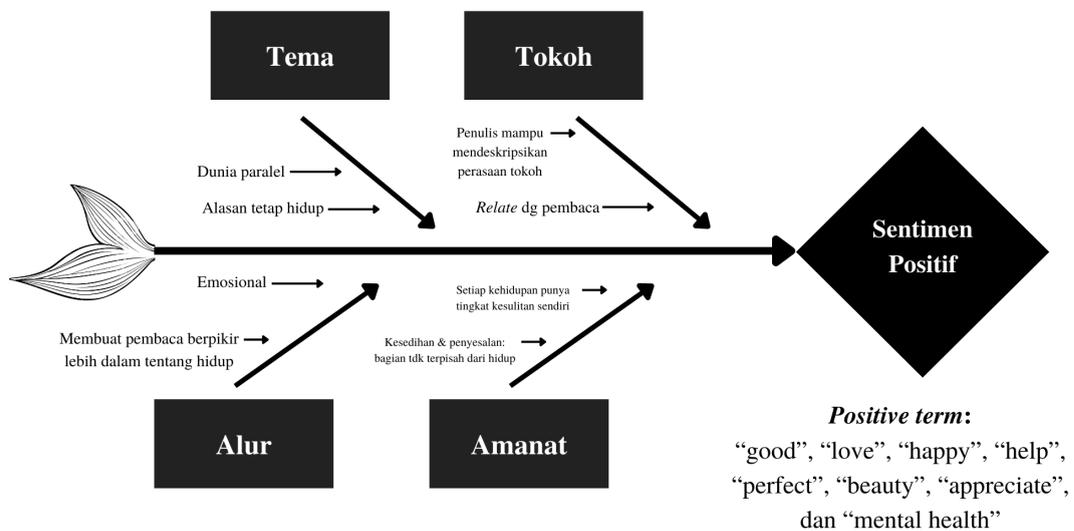
3. Alur cerita

Pembaca merasa bosan dengan alur cerita yang dapat terus tertebak saat

membalikkan halaman novel. Sebagian besar cerita hanya menggambarkan Nora Seed yang terus bolak-balik antara perpustakaan tengah malam dan kehidupan lain yang ingin dicobanya. Alur ini berjalan terus di sepanjang alur cerita tanpa menunjukkan pengembangan karakter secara signifikan. Nora terkesan menjadi tokoh *clueless* yang membosankan. Selain itu, ada juga *reviewer* yang menulis bahwa dia tertarik membaca novel ini karena sedang *hype* dan banyak yang memberi komentar positif, tetapi dia kecewa karena merasa cerita novel biasa saja.

4. Amanat

Pada bagian ini, pembaca berharap adanya cerita mengenai kontemplasi perenungan pikiran Nora saat di perpustakaan tengah malam yang membuatnya berada di antara hidup dan mati. Terlebih lagi, selama di perpustakaan pembaca ingin ada lebih banyak dialog yang kuat dengan Ms. Elm yang biasa memberi nasihat-nasihat kehidupan untuk Nora. Selain itu, pembaca juga merasa pesan yang mengiringi keputusan Nora kembali ke kehidupan lamanya kurang *impactful*, malah seperti membaca tulisan-tulisan motivasi di sosial media saja.



Gambar 4.15 Diagram *Fishbone* Sentimen Positif

Diagram *fishbone* untuk sentimen positif di atas menyoroti pendapat pembaca mengenai tema, tokoh, alur, dan amanat yang terkandung dalam novel *The Midnight Library* karya Matt Haig.

1. Tema

Pembaca menyukai novel ini karena menceritakan tentang adanya dunia paralel (*parallel universe*), yakni kehidupan lain yang saat ini berjalan beriringan dengan kehidupan yang tengah dijalani seseorang. Selain itu, pembaca juga merasa diberi harapan dan motivasi untuk tetap kuat menjalani kehidupan dari kata-kata yang tertulis di novel ini seiring dengan perjalanan yang dialami Nora.

2. Tokoh

Pembaca merasa penulis dapat memberi gambaran karakter tokoh dengan baik dalam sosok Nora sebagai seorang yang depresi tetapi dapat bangkit kembali. Bahkan, pembaca merasa cerita yang dialami Nora sangat relevan dengan keadaan yang telah atau sedang mereka jalani.

3. Alur cerita

Pembaca merasa alur cerita sangat emosional dengan melibatkan kegelisahan dan perasaan tidak nyaman yang Nora rasakan. Apalagi di akhir cerita Nora dapat bangkit kembali. Alur ini sekaligus menjadi penyemangat juga bagi pembaca agar lebih bergairah menjalani kehidupan terlepas dari segala ketidaknyamanan yang telah dialami.

4. Amanat

Pembaca dapat mengambil pesan yang disampaikan pada novel ini bahwa setiap kehidupan pasti mempunyai kesulitannya sendiri. Hal ini direpresentasikan dari perilaku Nora Seed yang berulang kali mengubah jalan hidupnya, tetapi tetap saja merasa ada hal yang membuatnya susah dan tidak nyaman. Begitu juga kesedihan dan penyesalan yang merupakan bagian tak terpisahkan dari kehidupan. Hal-hal seperti ini harus seseorang jalani dan tidak bisa terhindar darinya.

Hasil dari diagram *fishbone* positif dan negatif diharapkan dapat digunakan pustakawan sebagai salah satu pertimbangan jika ingin melakukan pengadaan koleksi bertema serupa di perpustakaan. Adanya visualisasi ini bertujuan memudahkan pustakawan mengidentifikasi apakah novel *The Midnight Library* cocok menjadi salah satu koleksi berdasarkan karakteristik perpustakaan dan pemustaka yang kerap berkunjung ke perpustakaan.

Jika faktor-faktor yang menjadi keluhan atau pandangan dari review pembaca dirasa berat diterima target pemustaka, maka bisa jadi produk ini tidak sesuai dengan target market dan target audiens-nya. Demikian juga untuk koleksi-koleksi novel lain yang ingin dibeli namun harganya mahal sedangkan

anggaran pengadaan koleksi sangat terbatas. Harapannya dengan mengetahui pandangan-pandangan dari review pembaca sebelumnya, pustakawan dapat menjadikannya sebagai salah satu pertimbangan apakah novel ini layak dibeli ataukah lebih baik menambah koleksi lain saja.

Dengan demikian, jika perpustakaan ingin melaksanakan pengadaan novel untuk pengembangan koleksi, pustakawan dapat lebih dulu melakukan analisis sentimen untuk mengetahui pandangan pembaca terhadap novel tersebut dengan hasil yang cepat sekaligus akurat. Data hasil analisis yang dilakukan dapat menjadi bahan pertimbangan untuk *data-driven decision* (pengambilan keputusan berdasarkan *insight* yang didapat dari data). Tujuannya adalah memastikan anggaran dana yang dipergunakan perpustakaan untuk pengadaan koleksi benar-benar terpakai untuk koleksi yang sekiranya mendapat sambutan baik dari pembaca dengan memiliki banyak sentimen review novel yang bernilai positif.

Selain itu, analisis sentimen juga dapat diterapkan saat ada penulis yang menerbitkan novel baru, dan penerbit ingin tahu tanggapan masyarakat pembaca terhadap novel itu. Analisis sentimen dapat berisi pendapat pembaca terhadap penulis, tokoh, cara menyampaikan alur, hingga topik bahasan atau *genre* novel. Hasil analisis sentimen dapat membantu industri penerbitan membuat keputusan terbaik pada penerbitan novel selanjutnya agar novel yang diterbitkan mendapat sambutan yang baik di pasar dan mendapat jumlah penjualan yang tinggi.

Text mining yang mengubah teks tidak terstruktur menjadi format terstruktur bermaksud mengidentifikasi pola yang bermakna dan wawasan baru. Dalam era disrupsi teknologi, perpustakaan tradisional sedang kesulitan karena

penggunaannya yang semakin berkurang. Pustakawan harus bertindak cepat menghadapi kesulitan ini dengan memahami secara efektif keberadaan *big data* dan cara mengelolanya agar profesi ini tetap relevan. Misalnya dengan menerapkan layanan perpustakaan berbasis *big data* saat membuat rekomendasi bagi pemustaka.⁵³ Hal ini memungkinkan dengan mempelajari sains data (*data science*). Saat ini, data teks dapat dengan mudah ditemukan di ruang lingkup bidang perpustakaan maupun penerbitan bahan pustaka. Data tersebut dapat diolah melalui proses pembersihan data (*preprocessing*), dan analisis sebelum dilakukan visualisasi. Hal ini menjadi peluang yang kuat bagi para pustakawan profesional untuk menjadi lebih kompeten di era sekarang.

Sekarang juga dikenal sebutan, “pustakawan data” yang dapat mengembangkan dan mengintegrasikan layanan perpustakaan berbasis data. Akibatnya, perpustakaan untuk akademik dan penelitian telah meluas. Pustakawan dapat mengembangkan keterampilan, pengetahuan, dan keahlian lingkup kepustakawanan yang dinamis untuk memajukan layanan teknis kepustakawanan berbasis data.⁵⁴ Berkaitan dengan hal ini, sebelumnya juga telah terdapat kajian berkaitan dengan ilmu perpustakaan dan sains data selama masa pandemi yang berada di garis depan dan saling bekerjasama di era mis informasi.⁵⁵

⁵³ Adetayo, Adebowale Jeremy, Pauline Oghenekaro Adeniran, and Arinola Oluwatoyin Gbotosho. "Augmenting Traditional Library Services: Role of Smart Library Technologies and Big Data." *Library Philosophy and Practice* 6164 (2021): 1-15.

⁵⁴ Ashiq, Murtaza, and Nosheen Fatima Warraich. "A Systematized Review on Data Librarianship Literature: Current Services, Challenges, Skills, and Motivational Factors." *Journal of Librarianship and Information Science* (2022): 09610006221083675.

⁵⁵ Tari, Mina, Yvette Iribe Ramirez, and Stephanie Ballard. "A Pandemic Makes for Strange Bedfellows: How Misinformation is Best Fought through Library & Information Science, Data Science, and Critical Theory." The Information School, University of Washington, 2020.

Perpustakaan dapat memanfaatkan sejumlah besar data yang tersimpan untuk peningkatan mutu layanan perpustakaan. Sumber data perpustakaan modern memiliki karakteristik *big data*, yakni perpustakaan dapat menggunakannya untuk mencapai reformasi dan inovasi pemanfaatan sumber daya demi meningkatkan pengalaman pemustaka.⁵⁶ Secara khusus, pustakawan akademik telah lama terlibat dalam pengelolaan data penelitian. Misalnya di Italia, pustakawan data adalah figur yang masih terus dikembangkan.⁵⁷

Pustakawan dapat memanfaatkan kesempatan ini dengan memiliki kemampuan manajemen data, bahasa mesin, dan analisis statistik. Keterampilan literasi data sangat penting untuk merevitalisasi peran pustakawan.⁵⁸ Sebelumnya juga terdapat penelitian mengenai pandangan dan kemampuan pustakawan untuk implementasi analisis *big data* di perpustakaan akademik Pakistan. Tujuannya adalah memeriksa hubungan antara keterampilan yang diperlukan pustakawan dan penerapan analisis *big data*. Penelitian ini menunjukkan bahwa terdapat korelasi yang kuat antara kompetensi dan keterampilan yang dibutuhkan pustakawan untuk implementasi analisis *big data* di perpustakaan akademik.⁵⁹

Konsep ini merupakan modernisasi teknologi dalam organisasi dan lembaga pendidikan untuk kemampuan analisis data berjumlah banyak yang

⁵⁶ Hamad, Faten, Hussam Fakhuri, and Sinaria Abdel Jabbar. "Big Data Opportunities and Challenges for Analytics Strategies in Jordanian Academic Libraries." *New Review of Academic Librarianship* 28, no. 1 (2022): 37-60.

⁵⁷ Morriello, Rossana. "Birth and Development of Data Librarianship." *JLIS.it: Italian Journal of Library, Archives, and Information Science* 11, no. 3 (2020).

⁵⁸ Jaiswal, Priyanka, and A. Singh Pratap. "Big Data Analytics: A Catalyst for Re-Energizing LIS Education". *Library Philosophy and Practice (e-journal)* 3867 (2020): 1-13.

⁵⁹ Ahmad, Khurshid, Zheng JianMing, and Muhammad Rafi. "An Analysis of Academic Librarians Competencies and Skills for Implementation of Big Data Analytics in Libraries: A Correlational Study." *Data Technologies and Applications* (2019).

berguna di perpustakaan akademik. Pemanfaatan *big data* secara signifikan dapat dirasakan dengan kemudahan akses pemanfaatan perpustakaan akademik.⁶⁰

Saat ini data disimpan dalam kondisi selalu aktif, dan dapat diakses secara global oleh pengguna kapanpun dan dimanapun. Layanan dan arsip data di Inggris (*Data Archives Services*), dan di Amerika Serikat dan Kanada (Layanan Perpustakaan Data), merupakan faktor kunci munculnya kepastakawanan data. Fokus kepastakawanan data saat ini adalah penciptaan layanan perpustakaan yang baru. Pustakawan data memerhatikan layanan manajemen data dan kompleksitasnya dapat membantu identifikasi keterampilan teknis yang dibutuhkan. Pustakawan data tidak perlu menjadi sangat ahli sebagai pemrogram, ahli statistik, atau pengelola basis data, tetapi harus tertarik mempelajari bahasa tersebut, logika pemrograman, *database*, dan alat pencarian informasi.⁶¹

Selanjutnya juga terdapat penelitian yang mengulas pengembangan sains data (*data science*) sebagai disiplin yang berkaitan dengan dengan bias data dan etika, disertai peran ilmu informasi untuk mengatasi masalah ini. Peneliti ilmu informasi mempunyai banyak hal untuk ditawarkan pada sains data, hal ini karena latar belakangnya adalah keilmuan transdisiplin yang berorientasi pada manusia dan perilaku sosial. Oleh karena itu, sains data dalam ilmu informasi akan terus berkembang dalam beberapa dekade mendatang.⁶²

⁶⁰ Islam, AYM Atiquil, Khurshid Ahmad, Muhammad Rafi, and Zheng JianMing. "Performance-Based Evaluation of Academic Libraries in the Big Data Era." *Journal of Information Science* 47, no. 4 (2021): 458-471.

⁶¹ Semeler, Alexandre Ribas, Adilson Luiz Pinto, and Helen Beatriz Frota Rozados. "Data Science in Data Librarianship: Core Competencies of a Data Librarian." *Journal of Librarianship and Information Science* 51, no. 3 (2019): 771-780.

⁶² Lund, Brady, and Ting Wang. "What Does Information Science Offer for Data Science Research?: A Review of Data and Information Ethics Literature." *Journal of Data and Information Science* (2022).

Masyarakat sudah mulai bertransisi dari era informasi menuju era data yang bervolume besar. Pustakawan data pada dasarnya telah menjadi jabatan dan posisi pekerjaan teraktual di bidang perpustakaan, namun American Library Association (ALA) belum mengembangkan kerangka inti kompetensi untuk profesi ini sebagai spesialisasi kurikulum untuk bidang yang tumbuh dan berkembang pesat.⁶³ Namun demikian, saat ini telah tersedia pembelajaran formal jika ingin menjadi pustakawan data di Amerika Serikat. Misalnya dengan mengikuti program Master of Library Science dari Luddy School of Informatics, Computing & Engineering, Indiana University. Program ini menyediakan pilihan konsentrasi sains data untuk mahasiswanya demi menunjang ilmu perpustakaan sebagai kajian di bidang data dan informasi.

Akhirnya, keberadaan inovasi kurikulum baru akan membantu persiapan mahasiswa ilmu perpustakaan menjadi seorang profesional yang berwawasan di bidang data. Pengembang kurikulum bisa mengoptimalkan manfaat dari studi ini dengan berfokus untuk mempersiapkan mahasiswa ilmu perpustakaan menjadi posisi pustakawan data di perpustakaan abad ke-21.⁶⁴ Mengesampingkan perbedaannya dengan ilmu informasi, sains data (*data science*) dapat digunakan sebagai metode baru untuk diaplikasikan pada layanan perpustakaan berbasis data. Terutama jika pustakawan mau berkenalan lebih jauh dengan logika dan bahasa pemrograman, serta visualisasi data.⁶⁵

⁶³ Khan, Hammad Rauf, and Yunfei Du. "What is a Data Librarian?: A Content Analysis of Job Advertisements for Data Librarians in The United States Academic Libraries." *International Federation of Library Associations and Institutions (IFLA)*, 2017.

⁶⁴ Khan, Hammad. "Exploring Data Science Learning Objectives in LIS Education." *Data Curation*. Ideals, University of Illinois, 2020.

⁶⁵ Koltay, Tibor. "Research Data Management and Data Literacy as We See Them Today." *Library And Information Science XXVIII* (2019): 7-17.