BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Penelitian ini menggunakan data media sosial berisi opini masyarakat dari bank digital yang ada di Indonesia sebagai objek penelitian. Data media sosial yang digunakan adalah data *tweet* dan data komentar pada konten akun Instagram resmi dari setiap bank digital. Data *tweet* dan komentar yang berisi opini dari masyarakat umum dan nasabah kemudian akan dianalisis sentimen dan topik utamanya mengenai bank digital tertentu.

Sampai bulan Juni 2021, terdapat delapan bank digital di Indonesia yaitu Jenius, Digibank, TMRW, Jago, Blu BCA Digital, Neobank, Seabank, dan Line Bank [5] [36]. Berdasarkan percobaan *scraping data* dalam tujuh hari pertama pengumpulannya yaitu dari tanggal 6 sampai 12 November, diperoleh tiga bank digital dengan ketersediaan data yang paling banyak di atas 500 data, yaitu Jenius, Jago, dan Line Bank. Hasil pengambilan data setiap bank terdapat pada tabel 3.1.

Tabel 3.1 Jumlah Data Pemilihan Objek Penelitian

Bank	Data Twitter	Data Instagram	Total Data
Line Bank	807	72	**879**
Jago	518	216	**734**
Jenius	399	144	**543**
Seabank	31	198	229
Blu	186	36	222
Digibank	95	108	203
Neobank	14	162	176
TMRW	19	108	127

3.1.1 Bank Jenius

Jenius merupakan layanan perbankan digital buatan Bank Tabungan Pensiunan Nasional (BTPN) yang resmi beroperasi di Indonesia sejak tanggal 11 Agustus 2016 [37]. Jenius menjadi produk bank digital pertama di Indonesia yang dapat diakses menggunakan *smartphone*. Jenius memiliki tiga akun media sosial resmi yaitu *@JeniusConnect* untuk Instagram dan Twitter, serta "Jenius Connect" pada Facebook.

Jenius memiliki beberapa fitur untuk tabungan (*save it)* untuk berbagai keperluan nasabah seperti *flexi*, *dream*, dan *maxi saver*. Fitur lainnya yang ada yaitu fitur peminjaman *flexi cash*, pengiriman uang (*send it)*, pembayaran tagihan dan *e-wallet* (*pay me*), pembagian bayaran tagihan untuk sesama nasabah Jenius (*split bill*), katu elektronik (*e-card*), dan juga adanya *$cashtag* yang bekerja seperti *username* untuk metode tambahan saat ingin mengirimkan uang kepada sesama nasabah Jenius [38].

3.1.2 Bank Jago

Bank Jago merupakan aplikasi bank digital yang dibuat oleh Bank Artos serta Gojek dan mulai resmi beroperasi di Indonesia pada tanggal 15 April tahun 2021 [39]. Bank Jago memiliki dua akun resmi pada media sosial Instagram dan Twitter yaitu *@jadijago* dan *@tanyajago*. Akun Youtube resminya juga bernama "jadijago" serta "JadiJagoOfficial" pada Facebook.

Bank Jago memiliki banyak sub fitur yang cukup banyak, tetapi untuk fitur utamanya mencakup pembuatan nama JagoID yang bekerja seperti *username*, pemisahan tabungan serta kolaborasi dalam menabung (*pockets*), pengiriman uang (*pay and send*), permintaan uang untuk sesama nasabah Jago (*request money*), pembayaran tagihan dan *e-wallet*, analisis pengeluaran (*spend analysis*), pembuatan kartu debit, serta yang terbaru adalah investasi keuangan yang terintegrasi dengan aplikasi Bibit [40].

3.1.3 Line Bank

Line Bank merupakan layanan perbankan digital milik Bank KEB Hana Indonesia bersama dengan Line Financial Asia yang mulai dirilis di Indonesia pada 10 Juni tahun 2021 [6]. Sebelumnya Line Bank sudah terlebih dahulu beroperasi di Thailand dan Taiwan, sehingga Indonesia menjadi negara ketiga yang menjadi target pasar Line Bank [6]. Media sosial yang dimiliki Line Bank khusus di Indonesia adalah *@linebankid* di Instagram, Twitter, dan Tiktok, serta "LINE Bank by Hana Bank" di Facebook dan juga Youtube.

Line Bank memiliki beberapa fitur yang juga serupa dengan bank digital lainnya, seperti pembukaan rekening secara cepat menggunakan e-KYC (*electronic know your customer*), akun deposit, pembuatan kartu debit dengan gambar karakter maskot Line, pengiriman dan pengambilan uang tanpa biaya tambahan, serta pembayaran tagihan [6]. Sebagai produk dari Line, nasabah juga dapat menghubungkan akun Line Messenger dengan akun Line Bank untuk mendapatkan notifikasi melalui Line Messenger mengenai transaksi di rekeningnya [6].

3.2 Metode Penelitian

3.2.1 Metode Penyelesaian

Penelitian ini akan menganalisis sentimen publik berdasarkan data media sosial Twitter dan Instagram dari tiga bank digital di Indonesia, serta melakukan *topic modelling* untuk menemukan topiknya. Menurut dua artikel *literature review* mengenai analisis sentimen [41], [42], metode *machine learning* yang paling sering digunakan untuk kasus analisis sentimen data media sosial adalah Naïve Bayes dan Support Vector Machine (SVM). Berikut adalah perbandingan antara kedua algoritma yang paling sering digunakan tersebut pada tabel 3.2:

Tabel 3.2 Tabel Perbandingan Algoritma NB dan SVM

Naïve Bayes (NB)	**Support Vector Machine (SVM)**
Supervised learning [28]	*Supervised learning* [28]
Berjenis *probabilistic classifier* [43]	Berjenis *linear classifier* [43]

Naïve Bayes (NB)	Support Vector Machine (SVM)
Atribut yang digunakan pada input tidak harus numerik. [28]	Semua atribut pada input harus berupa numerik. [28]
Mampu membuat model dengan jumlah *training data* yang sedikit. [44]	*Training data* yang diperlukan untuk membuat model harus berjumlah banyak. [44]
Pembuatan model hingga *deployment* memakan waktu yang lebih cepat. [28]	Membutuhkan waktu yang lebih lama untuk pembuatan model. [28]
Bekerja lebih baik pada dimensi input yang lebih sedikit [44], dengan asumsi semua atribut atau *feature* bersifat independen [28]	Mampu memproses input yang *high dimension*. [44]
Hubungan probabilistik antara setiap *feature* dengan target output (label) yang digunakan untuk klasifikasi. [28]	Mempelajari perwakilan dari keseluruhan data agar dapat terpisah setiap kategorinya dengan jarak yang maksimum. [29]

Pada tugas selanjutnya yaitu pemodelan topik atau *topic modelling*, algoritma yang paling sering digunakan menurut *literature review* adalah Latent Dirichlet Allocation (LDA) [17], [18] dan Latent Semantic Analysis (LSA) [18]. Berikut tabel 3.3 menjelaskan mengenai perbedaan antara kedua metode tersebut:

Tabel 3.3 Tabel Perbandingan Algoritma LDA dan LSA

Latent Dirichlet Allocation (LDA)	Latent Semantic Analysis (LSA)
Probabilistic topic modelling. [19]	*Linear topic modelling.* [19]
Jumlah topik harus ditentukan di awal penelitian, dapat ditentukan dengan menggunakan nilai *coherence.* [20]	Jumlah topik yang akan dihasilkan sulit untuk ditentukan, serta harus dilakukan secara manual. [20]
Mampu mengolah dokumen yang pendek, panjang, dan juga *mixed-length.* [20]	Memiliki kemampuan yang kurang dalam mengolah dokumen yang pendek. [20]
Hubungan antar dokumen ikut dipertimbangkan. [19]	Hubungan antar dokumen tidak menjadi hal yang dipertimbangkan. [19]

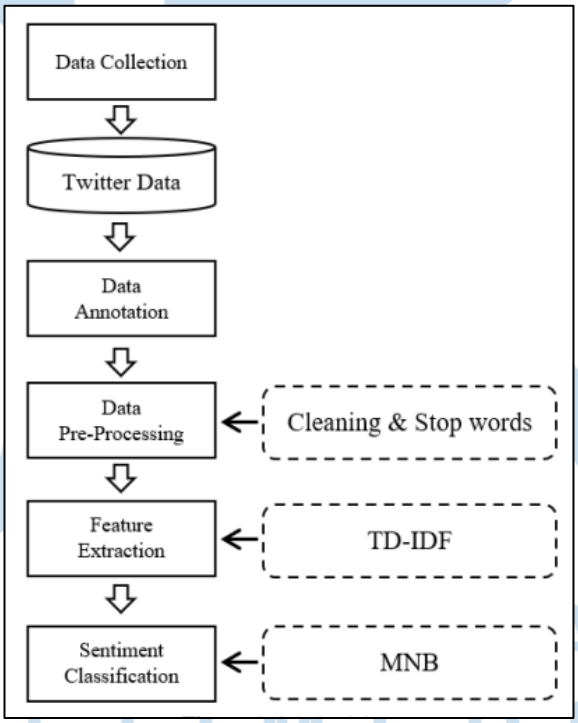
Berdasarkan hasil perbandingan algoritma di tabel 3.1, algoritma Naïve Bayes dipilih untuk melakukan tugas klasifikasi sentimen secara otomatis dari media sosial setiap bank digital. Pemilihan ini juga didukung dengan beberapa penelitian yang membuktikan keunggulan performa Naïve Bayes pada klasifikasi sentimen, jika dibandingkan dengan algoritma lainnya pada setiap penelitian tersebut [22] - [25].

Pada tugas *topic modelling* untuk menentukan topik utama dari opini publik atas setiap bank digital, maka dipilih algoritma Latent Dirichlet Allocation sebagaimana dibuktikan keunggulannya pada penelitian [19], [20], [21]. Topik akan ditentukan pada setiap kategori sentimen (positif dan negatif), dimana topik pada kategori positif bisa menandakan kelebihan atau

keunggulan dari bank digital tertentu yang harus dipertahankan performanya. Hasil topik kategori negatif dapat digunakan untuk bahan evaluasi setiap bank digital mengenai topik tertentu yang akan didapatkan.

3.2.2 Alur Penelitian

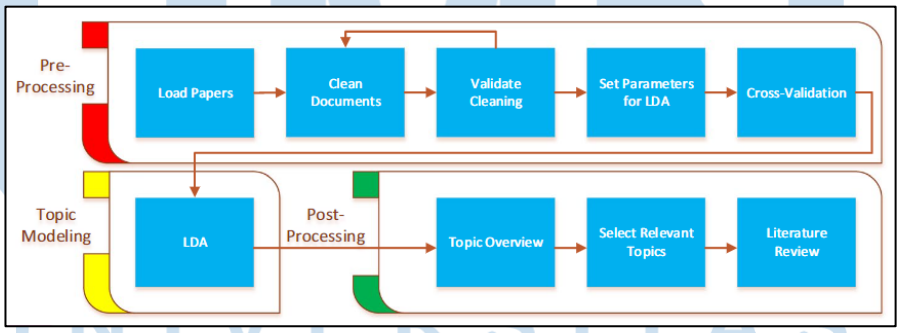
Pada penelitian ini yang bertemakan analisis sentimen, alur penelitian yang dilakukan akan berjalan seperti alur penelitian analisis sentimen pada umumnya namun ditambahkan langkah tambahan untuk pemodelan topik dari setiap bank digital. Contoh alur penelitian yang dipilih sebagai referensi mengenai analisis sentimen dengan Naive Bayes berasal dari referensi artikel jurnal berjudul *Philippine Twitter Sentiments during Covid-19 Pandemic using Multinomial Naïve-Bayes* [45]. Berikut alur penelitian dari artikel jurnal tersebut pada gambar 3.1:



Gambar 3.1 Alur Penelitian Terdahulu: Analisis Sentimen [45]

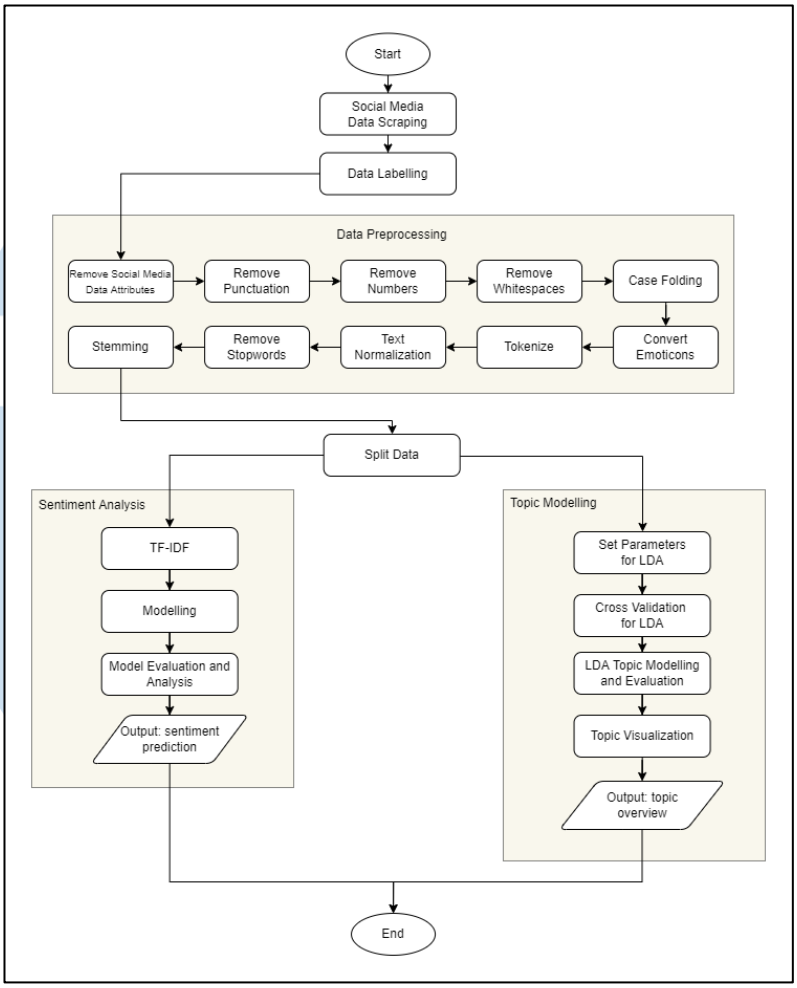
Terdapat beberapa modifikasi langkah yang akan dilakukan dalam penelitian ini jika dibandingkan dengan alur pada gambar 3.1. Perbedaannya berada pada tahap pengumpulan data yang bersumber dari dua media sosial, bukan hanya satu media sosial. Pada penelitian tersebut hanya dilakukan 4 langkah *pre-processing* yaitu *remove social media attribute, remove punctuation, stopwords,* serta langkah *case folding.* Oleh karena itu tahap data *pre-processing* di penelitian ini akan disesuaikan dan ditambah sesuai data dari dua media sosial yang diperoleh. Penelitian ini juga tidak berhenti sampai di tahap klasifikasi sentimen, tetapi selanjutnya akan ditambah dengan tugas pemodelan topik.

Alur proses pemodelan topik akan menggunakan referensi kerangka kerja pada artikel jurnal penelitian terdahulu berjudul *Smart literature review: a practical topic modelling approach to exploratory literature review* [46]. Terdapat perubahan pada langkah permulaannya dikarenakan tahap *preprocessing* perlu dilakukan untuk kedua proses analisis sentimen dan juga pemodelan topik. Tahap *clean document* akan digabung dengan tahap *preprocessing* sebelum analisis sentimen untuk efisiensi.. Oleh karena itu, alur akan dimulai setelah langkah *preprocessing* selesai. Berikut gambar 3.2 adalah kerangka kerja pada artikel tersebut:



Gambar 3.2 Alur Penelitian Terdahulu: Pemodelan Topik [46]

Berdasarkan kedua alur penelitian terdahulu mengenai analisis sentimen dan pemodelan topik, berikut gambar 3.3 adalah alur penelitian hasil modifikasi penulis untuk penelitian ini:



Gambar 3.3 Alur Penelitian Hasil Modifikasi

Seluruh proses yang dilakukan nantinya juga akan diulang sebanyak tiga kali, untuk mendapatkan hasil analisis sentimen dan pemodelan topik untuk tiga bank digital. Penjelasan mengenai setiap langkah yang dilakukan di gambar 3.3 adalah sebagai berikut:

1. *Social Media Data Scraping:*

   Tahap *scraping data* dilakukan pada dua media sosial yaitu Twitter dan Instagram untuk setiap bank digitalnya. Pada media sosial Twitter, data diambil menggunakan *library tweepy* pada *Python.* Agar dapat menggunakan *library* ini perlu melakukan permintaan akses *Application Programming Interface* (API) Twitter terlebih dahulu pada situs Twitter Developer.

Setelah mendapatkan akses, data *tweet* akan diambil berdasarkan kata kunci nama bank, serta *username* dari akun resmi yang dimiliki bank. Kata kunci untuk Bank Jenius adalah "bank jenius" dan "@jeniusconnect", untuk Bank Jago yaitu "bank jago" dan "@jadijago", serta untuk Line Bank yaitu "line bank" dan "@linebankid".

Media sosial kedua yang menjadi sumber data adalah Instagram, dimana data akan diambil pada komentar dari konten yang diunggah di *feeds* milik akun resmi setiap bank digital. Akun Instagram resmi yang menjadi tujuan adalah *@jeniusconnect @jadijago* dan *@linebankid*.

2. *Data Labelling*

Pada tahap ini akan dilakukan pemberian label sentimen (positif atau negatif) pada setiap data *tweet* atau komentar Instagram. Langkah ini perlu dilakukan untuk dapat membangun model prediksi klasifikasi yang akan dilakukan dengan algoritma berbasis *supervised learning.* Pemberian label sentimen akan dilakukan dengan cara manual seperti pada penelitian terdahulu mengenai analisis sentimen dan juga pemodelan topik [22][45][47]. Jumlah pemberi label sebanyak tiga orang tidak termasuk penulis. Penentuan hasil akhir sentimen akan dilakukan dengan memilih hasil *majority voting,* dimana label dengan jumlah pemilih terbanyak akan dipilih sebagai label sentimen tersebut.

3. *Remove Social Media Data Attributes*

Atribut khusus dapat ditemukan pada data media sosial, seperti adanya *hashtag '#', mention '@', link,* serta spesifik pada data Twitter terdapat penanda 'RT' untuk jenis data yang di *retweet.* Untuk meningkatkan kualitas data, seluruh atribut ini akan dihapuskan dari datanya [45].

4. *Remove Punctuation*

Tanda baca dianggap tidak memiliki atau mewakili sebuah makna, kecuali untuk menghubungkannya dengan tata bahasa [48]. Pada data media sosial dapat ditemukan tanda baca yang menandakan fungsi berbeda seperti '@' untuk menandakan *username* dan '#' untuk menandakan *hashtag*. Tanda baca lainnya juga sering ditemukan pada *link* sebuah *website.* Oleh karena itu, tanda baca tidak akan digunakan untuk proses analisis. Tanda baca yang dihapus diambil dari *library 'string'* pada *Python*, yaitu !"#$%&'()*+, -./:;<=>?@[\]^_`{|}~ [49].

5. *Remove Numbers*

Selain menghapuskan tanda baca, seluruh digit angka yang ada pada data media sosial juga akan dihapuskan untuk mencegah *noisy data* dikarenakan angka yang muncul tidak memiliki makna untuk penentuan sentimennya [50].

6. *Case Folding*

Untuk mencegah perhitungan kata yang sama tetapi memiliki kapitalisasi yang berbeda, tahap *case folding* dilakukan dengan mengubah semua huruf menjadi non kapital [51].

7. *Remove Whitespaces*

Untuk mengurangi banyaknya data yang akan diproses, penghapusan *whitespace* dilakukan karena tidak memiliki makna apapun yang dapat memengaruhi data teks [52].

8. *Convert Emoji*

Pada artikel yang digunakan sebagai referensi alur penelitian, perubahan hanya dibagi menjadi dua jenis, yaitu kumpulan *emoji* yang diubah menjadi kata 'sedih' dan 'senang' [51]. Pada penelitian ini akan digunakan *library emot* untuk mengganti berbagai *emoji* menjadi kata-kata yang mewakilkan *emoji* tersebut.

9. *Tokenize*

Tahapan tokenisasi merupakan langkah untuk memecah kalimat yang ada menjadi satuan yang lebih kecil, dimana pada kasus data media sosial, akan dipecah menjadi kata per kata [53].

10. *Text Normalization*

Proses *normalization* bekerja dengan cara memetakan beberapa kata, misalnya kata atau singkatan yang tidak formal (*slang*), menjadi bentuk kata formal yang terstandarisasi.[54]. Kata-kata yang tidak formal banyak ditemukan pada data media sosial yang berisi ungkapan perasaan penggunanya. Oleh karena itu penelitian ini menggunakan *dictionary* sesuai standar KBBI yang kemudian dilakukan *mapping* sesuai *dictionary* tersebut.

11. *Remove Stop Words*

Stop words merupakan kumpulan kata-kata yang sangat sering ditemukan pada sebuah kalimat dan dapat membuat kata-kata lainnya yang lebih relevan menjadi tidak terhitung [51]. Pada penelitian ini akan digunakan daftar kata-kata *stop words* dalam bahasa Indonesia di *library nltk.corpus* untuk kemudian dihapuskan.

12. *Stemming*

Setiap kata yang ada akan diubah menjadi kata dasarnya berdasarkan kesesuaian dengan aturan bahasa Indonesia [51]. Tahap *stemming* akan dilakukan menggunakan *library Python* yaitu *sastrawi.*

13. *Split Data*

Pemisahan data bertujuan agar membagi data menjadi data *training* untuk membangun model klasifikasi, validasi, serta *testing* untuk menguji model yang sudah dibangun. Pembagian data akan dilakukan secara sama

rata untuk setiap bank. Pemisahan data pertama dimulai dari memisahkan 80% data setiap media sosial untuk kebutuhan pembuatan model dan validasinya, kemudian sebanyak 20% dari data untuk kebutuhan *testing.*

Selanjutnya 80% data dari setiap media sosial akan digabungkan untuk membangun model klasifikasinya. Berdasarkan gabungan data media sosial tersebut akan dipisahkan lagi menjadi 80% data untuk training, dan 20% data untuk validasi.

Pemisahan data untuk kebutuhan tugas kedua yaitu pemodelan topik akan dilakukan per kategori sentimen untuk setiap *dataset* bank digital. Sehingga pemodelan topik akan dilakukan dua kali, pada data yang bersentimen positif dan negatif, untuk setiap *dataset* bank digital.

14. Create TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan metode pembobotan kata dari *dataset* yang akan dilakukan pada penelitian ini. Pembuatan TF-IDF perlu dilakukan untuk mengubah data teks menjadi matriks kata dan bobotnya agar dapat diolah oleh algoritma yang digunakan [55]. Kata yang lebih jarang muncul akan mendapat nilai yang lebih tinggi dan dianggap lebih relevan, serta kata yang terlalu sering muncul akan dikurangi karena dianggap kurang relevan [56]. TF-IDF terbagi menjadi dua proses, yaitu perhitungan frekuensi kata pada sebuah dokumen dan dibagi dengan jumlah kata pada dokumen (TF), kemudian tahap IDF adalah kata diberikan bobot berdasarkan perhitungan logaritma banyaknya dokumen dibagi dengan banyaknya dokumen yang mengandung kata tersebut [57].

15. Naïve Bayes Classification Modelling

Proses *modelling* di penelitian ini akan dimulai dari pemilihan parameter terbaik untuk modelnya menggunakan *function GridSearchCV*, dengan menggunakan parameter *cross validation* untuk meningkatkan kinerjanya. Setelah mendapatkan parameter terbaik model akan dibangun

berdasarkan *data training,* yang berupa gabungan data kedua media sosial.

Model yang sudah dibangun akan digunakan untuk memprediksi data *testing* yang dibagi menjadi per media sosial. Tujuannya adalah untuk membandingkan performa model pada data baru per media sosial yang tidak menjadi bagian data *training* dan validasi. Seluruh proses *modelling* ini akan dilakukan untuk setiap dataset bank digital yang ada, yaitu Bank Jenius, Jago, dan Line Bank.

16. Model Evaluation and Analysis

Untuk penentuan kinerja model klasifikasi utamanya akan digunakan nilai F1 *score* agar dapat mengukur kinerja untuk jumlah data yang tidak seimbang per labelnya [58]. Sedangkan nilai lainnya seperti akurasi, presisi, dan *recall* akan digunakan untuk menganalisis kinerjanya berdasarkan *confusion matrix.* Proses klasifikasi sentimen berhenti pada langkah ini setelah diulang seluruh prosesnya untuk setiap data media sosial bank digital.

17. Set Parameters for LDA

Algoritma LDA membutuhkan parameter sebelum memulai tahap pemodelan, yaitu parameter jumlah optimal dari topik yang akan dihasilkan. Semakin banyak jumlah topik biasanya menandakan topik yang semakin rinci, sedangkan jumlah topik yang sedikit menandakan topik yang semakin umum [46]. Selain itu parameter yang disiapkan adalah data dalam bentuk *dictionary, corpus,* serta *bag of words* yang mempertimbangkan *n-grams*.

18. Cross Validation for LDA

Tahap *cross validation* dilakukan untuk memastikan pemodelan yang dilakukan dengan algoritma LDA menghasilkan jumlah topik yang optimal [46]. Proses validasi jumlah topik optimal akan menggunakan

nilai *coherence*, dimana tingkat kemiripan antara kata-kata yang tergabung dalam satu topik adalah hal yang digambarkan pada nilai ini [59]. Nilai *coherence* yang semakin besar menandakan topik yang dihasilkan semakin baik dalam mewakili dokumen [60].

19. LDA Topic Modelling and Evaluation

Pembuatan model pemodelan topik dengan LDA dilakukan dengan menggunakan parameter terbaik yang didapat dari tahapan sebelumnya, kemudian diimplementasikan pada data media sosial yang sudah melalui tahap *preprocessing.* Penentuan topik akan dilakukan pada data per label sentimen, sehingga dapat diketahui topik utama pada data bersentimen positif dan negatif untuk setiap bank digital.

20. Topic Visualization

Hasil proses pemodelan topik akan divisualisasikan dengan menampilkan daftar topik beserta dengan grafik menggunakan *library Python pyLDAVis.* Tampilan utama yang ada pada grafik adalah *intertopic distance map* yang dapat digunakan secara interaktif atau tidak statis [61]. Jika visualisasi *cluster map* ditekan salah satu topiknya, maka *bar chart* akan berubah sesuai dengan frekuensi kata pada topik terpilih tersebut. Visualisasi menggunakan pyLDAVis dapat menambah pemahaman mengenai hasil topik yang dihasilkan [61].

3.3 Variabel Penelitian

3.3.1 Variabel Independen

Variabel independen atau variabel bebas yang digunakan pada penelitian ini adalah opini masyarakat mengenai bank digital Jenius, Jago, dan Line Bank yang berada di media sosial Twitter dan Instagram. Data dari Twitter adalah data *tweet* dengan kata kunci tertentu, sedangkan data Instagram adalah kumpulan komentar yang diambil dari konten akun resmi

setiap bank digital. Ketentuan rinci dari setiap pengambilan data tersebut terdapat pada bab 3.2.2.

3.3.2 Variabel Dependen

Variabel dependen atau variabel terikat pada penelitian ini adalah hasil label sentimen dari setiap data media sosial bank digital. Kategori atau label sentimen yang digunakan ada dua, yaitu sentimen positif dan negatif. Selain label sentimen, topik yang dihasilkan pada tahap *topic modelling* juga merupakan variabel dependen.

3.4 Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data primer, yang diperoleh dengan cara *scraping* dari setiap media sosial menggunakan bahasa pemrograman Python di Jupyter. *Library* Python yang digunakan untuk *scraping* data *tweet* adalah *tweepy* untuk mengakses API Twitter [62]. Data komentar dari Instagram didapatkan dengan *library* Python *Instagram-Comments-Scraper* [63].

Data media sosial Twitter dan Instagram diambil dalam periode waktu tiga bulan sejak 6 November 2021 hingga 19 Maret 2022. Penentuan periode waktu ditentukan berdasarkan bank digital yang paling baru dirilis yaitu Line Bank pada pertengahan tahun 2021. Berdasarkan ketentuan API Twitter periode untuk *scraping* juga diberikan batas selama tujuh hari sebelum tanggal *scraping*, sehingga data Twitter akan diambil secara berkala setiap minggunya.

3.5 Teknik Pengambilan Sample

Teknik pengambilan sample data yang dilakukan pada penelitian ini adalah dengan cara menguji pengambilan sample data media sosial yang diperoleh dalam waktu satu minggu. Pemilihan waktu satu minggu untuk pengambilan *sample* dipilih berdasarkan ketentuan API Twitter yang hanya membolehkan pengambilan data dalam batas tujuh hari sebelum hari dilakukannya *scraping*. Data ini kemudian

akan diperiksa oleh tiga orang selain penulis, dimana hasil pengujiannya akan ditentukan berdasarkan hasil *majority vote*. Hal yang diuji terlebih dahulu pada kedua data media sosial adalah mengenai pemberian label sentimennya.

3.6 Teknik Analisis Data

Tersedia banyak *tools* yang dapat digunakan untuk melakukan berbagai tugas analisis data. Pada hasil pemungutan suara yang dilakukan KDNuggets pada tahun 2019 berjudul *Top Analytics, Data Science, Machine Learning Software*, Python dan RapidMiner menduduki peringkat 1 dan 2 secara berurut diantara 90 pilihan *tools* lainnya sejak tahun 2017 [64]. Salah satu Integrated Development Environment (IDE) dari Python yang dibandingkan performanya pada beberapa penelitian terdahulu mengenai *tools* untuk *data analytics*, *data mining*, dan *text preprocessing* adalah Jupyter [48], [57], [65], [66]. Oleh karena itu, Jupyter dan RapidMiner akan dibandingkan pada tabel 3.4 berikut:

Tabel 3.4 Tabel Perbandingan *Tools*

Jupyter	RapidMiner
Dibangun dengan bahasa pemrograman C dan Python [65].	Dibangun dengan bahasa pemrograman Java [65].
Code, hasil output, dan visualisasi dapat ditampilkan secara langsung pada *notebook* [65].	Platform berbasis GUI (graphical user interface) dengan konsep *drag and drop* [67].
Tidak terbatas untuk tugas *machine learning*, dapat berfungsi untuk pemrograman secara umum [65].	Berfungsi untuk *machine learning, predictive analytics*, dan *business intelligence* [65].
Memiliki tampilan layout yang sudah terstandarisasi [68].	Layout menampilkan proses kerja secara intuitif [68].
Kurangnya dukungan komersial [65].	Merupakan *tools* yang komersial [65].

Dengan mempertimbangkan faktor-faktor pada tabel 3.3 di atas, penelitian ini akan menggunakan Jupyter dengan bahasa pemrograman Python dalam menganalisis sentimen dan melakukan pemodelan topik dari pada data media sosial dari tiga bank digital di Indonesia. Salah satu faktor lainnya yang mendukung pemilihan Python adalah adanya *library 'pyLDAVis'* yang dibutuhkan pada penelitian ini untuk visualisasi data secara interaktif dari hasil pemodelan topik menggunakan algoritma Latent Dirichlet Allocation.