

Analisis Pinjaman Online Pada Sosial Media *Twitter* Menggunakan *Latent Dirichlet Allocation (LDA)*

Puji Septiyana Nautika¹, Wiyli Yustanti²

^{1,2} Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya

puji.20036@mhs.unesa.ac.id

wiyliyustanti@unesa.ac.id

Abstrak— Industri keuangan Indonesia mengalami transformasi besar selama era digitalisasi dan kemajuan teknologi informasi. Pertumbuhan bisnis start-up fintech di Indonesia, khususnya melalui penggunaan aplikasi pinjaman online, membawa fenomena pinjaman online. Pinjaman online telah menimbulkan berbagai kontroversi dan isu yang menjadi sorotan dikalangan masyarakat. Diperlukan suatu metode yang cepat dan efisien untuk mengolah dan memahami informasi yang terdapat dari jutaan tweet agar lebih terstruktur dan sistematis yaitu dengan metode pemodelan topik. . LDA dapat digunakan untuk memodelkan topik-topik laten yang ada di dalam kumpulan dokumen. Pada penelitian ini, digunakan metodologi *Knowledge Discovery in Databases (KDD)* untuk mengidentifikasi dan mengelompokkan isu-isu dari tweet mengenai pinjaman online. Penelitian ini bertujuan untuk membantu dalam mengidentifikasi dan mengelompokkan isu-isu dari tweet mengenai pinjaman online dengan metode *Latent Dirichlet Allocation (LDA)*. Didapatkan pemodelan LDA terbaik pada skenario satu dengan parameter untuk iterasi bernilai 10, nilai α sebesar 0,1 dan β sebesar 0,1. Dalam pemodelan skenario satu ini didapat nilai coherence score tertinggi sebesar 0,391. Nilai ini lebih unggul dari skenario lainnya dengan nilai tertinggi dan didapat jumlah topik yaitu 9 topik.

Kata Kunci— Pemodelan Topik, Pinjaman Online, *Latent Dirichlet Allocation*, Coherence Score

I. PENDAHULUAN

Industri keuangan Indonesia mengalami transformasi besar selama era digitalisasi dan kemajuan teknologi informasi. Pertumbuhan bisnis start-up fintech (fintech) di Indonesia, khususnya melalui penggunaan aplikasi pinjaman online, membawa fenomena pinjaman online atau dikenal dengan istilah “pinjol”. Menurut data Otoritas Jasa Keuangan (OJK), fintech lending telah menyalurkan pinjaman online senilai Rp20,54 triliun pada Agustus 2023, naik 0,79% dari Rp20,38 triliun pada bulan sebelumnya. Ini adalah peningkatan sebesar 6,85% dibandingkan dengan data Agustus 2022, yang mencatat Rp19,22 triliun.

Untuk memberikan pinjaman, perusahaan rintisan menggunakan fintech atau teknologi finansial, yang memiliki syarat dan ketentuan yang lebih mudah dan fleksibel dibandingkan dengan institusi keuangan konvensional seperti bank. Masyarakat dapat mengakses layanan ini dengan nyaman dan mudah di rumah berkat penggunaan smartphone dan internet [1]

Aplikasi pinjol sangat populer di kalangan orang Indonesia karena sangat mudah digunakan, terutama bagi mereka yang

membutuhkan dana darurat atau tidak memiliki akses ke lembaga keuangan konvensional. Namun, di balik kemudahan yang ditawarkan oleh fintech atau pinjaman online, muncul berbagai masalah yang berdampak besar pada masyarakat. Suku bunga yang lebih tinggi dapat menyebabkan beban pinjaman yang lebih besar bagi peminjam, hal ini dapat menyulitkan untuk memenuhi kewajiban pembayaran mereka. Selain itu, banyaknya keberadaan layanan pinjaman online yang beroperasi tanpa izin resmi telah menimbulkan kekhawatiran di tengah masyarakat. Hal ini disebabkan oleh metode penagihan mereka yang seringkali mengganggu dan bersifat mengintimidasi [2], [3].

Kurangnya peraturan dan perlindungan hukum yang cukup untuk peminjam dana online menyebabkan berbagai fenomena hukum yang muncul dan mengganggu masyarakat. Selain itu, rendahnya pemahaman literasi keuangan di antara peminjam seringkali mengakibatkan mereka membuat keputusan finansial yang kurang tepat. pada Februari 2023, OJK menerima 636 aduan serupa, pada Maret 2023, 980 aduan, pada April 2023, 694 aduan, dan pada Mei 2023, 420 aduan [4]. Dengan adanya data tersebut, masyarakat dapat mewaspadai keputusan untuk melakukan pinjaman online dan membantu menekan angka literasi keuangan agar mengetahui topik mengenai kasus pinjaman online.

Pinjaman online telah menimbulkan berbagai kontroversi dan isu yang menjadi sorotan dikalangan masyarakat. Media sosial telah menjadi salah satu platform penting bagi masyarakat dalam era informasi digital untuk berkomunikasi, berbagi informasi, dan menyuarakan pendapat mereka. . Khususnya, Twitter, dengan jumlah pengguna aktif yang mencapai 327,9 juta pada April 2023 menurut laporan We Are Social [5]. Twitter telah menjadi salah satu platform media sosial yang digunakan untuk menyampaikan opini publik tentang berbagai masalah, termasuk masalah pinjaman online. Pengguna Twitter dapat berdiskusi melalui tweet yang membahas berbagai topik. Topik informasi dari berbagai tweet saling berkaitan.

Analisis data seperti pemodelan topik dan analisis sentimen dapat membantu menemukan informasi tersembunyi dari sekumpulan tweet [6]. Diperlukan suatu metode yang cepat dan efisien untuk mengolah dan memahami informasi yang terdapat dari jutaan tweet agar lebih terstruktur dan sistematis yaitu dengan metode pemodelan topik. Pemodelan topik adalah teknik untuk menemukan kelompok kata atau topik, dalam kumpulan dokumen. Kelompok kata ini dapat menunjukkan informasi yang ada dalam kumpulan dokumen

[7]. Metode Latent Dirichlet Allocation (LDA) menawarkan solusi dalam pemodelan topik. LDA dapat digunakan untuk memodelkan topik-topik laten yang ada di dalam kumpulan dokumen. Ini memberikan representasi probabilistik dokumen sebagai campuran topik, memungkinkan pemahaman dan pengorganisasian data tekstual yang lebih baik [8].

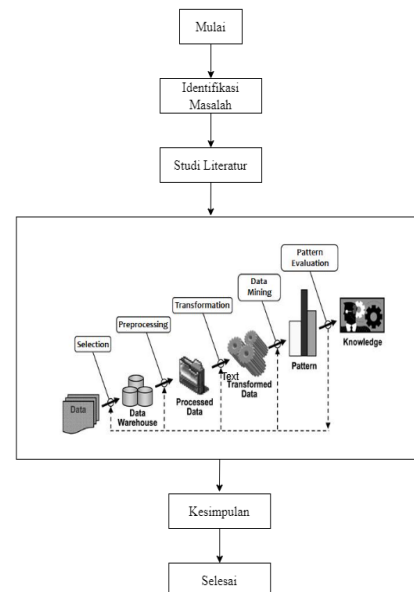
Latent Dirichlet Allocation (LDA) merupakan model probabilistik generatif dari korpus, dimana dokumen direpresentasikan sebagai kumpulan topik laten dengan distribusi kata-kata yang akan membedakan setiap topik [9]. LDA berasumsi bahwa setiap dokumen terdiri dari campuran topik yang berbeda yang masing-masing dapat dijelaskan melalui distribusi probabilitas. Metode Latent Dirichlet Allocation (LDA) untuk mengidentifikasi dan menggambarkan topik dalam kumpulan dokumen yang sangat besar dengan menganalisis kata-kata yang tersebar. LDA dipilih karena dapat menganalisis data dan dokumen yang sangat besar [10].

Penelitian terdahulu terkait pemodelan topik dengan metode Latent Dirichlet Allocation (LDA) yang dilakukan oleh [11], data yang digunakan berasal dari tweet pada aplikasi Twitter. Dari penelitian ini ditemukan 5 topik yaitu topik ke-1 mengenai penolakan UU Omnibus Law Cipta Kerja oleh sejumlah lapisan masyarakat dan kepala daerah, topik ke-2 yaitu kekhawatiran masyarakat terhadap penyebaran virus Corona yang disebabkan oleh pelanggaran kebijakan PSBB yang dilakukan oleh demonstran dalam unjuk rasa omnibus law, topik ke-3 mengenai dampak UU omnibus law cipta kerja terhadap kawasan hutan dan lingkungan, topik ke-4 mengenai keistimewaan yang didapat warga asing karena omnibus law, dan topik ke-5 kerusakan yang terjadi pada saat demo tolak omnibus law di berbagai wilayah. Ditemukan nilai dari *coherence score* sebesar 0.5644.

Penelitian saat ini menggunakan pemodelan topik *Latent Dirichlet Allocation* (LDA), yang mengelompokkan tweet masyarakat mengenai yang berfokus pada pemodelan topik terhadap isu pinjaman online yang relevan di platform Twitter. Menggunakan topik coherence sebagai metrik evaluasi untuk menentukan kelompok yang optimum. berfokus pada pemodelan topik terhadap isu pinjaman online dari Twitter, yang merupakan topik yang menarik dan penting untuk diteliti. Dengan demikian, penelitian ini dapat memberikan kontribusi baru pada pemahaman tentang topik yang dibahas dalam tweet tentang pinjaman online dan dapat memberikan wawasan baru bagi industri keuangan dan masyarakat umum.

II. METODOLOGI PENELITIAN

Metodologi penelitian berfungsi sebagai panduan dasar dalam pelaksanaan penelitian, memastikan penelitian dilakukan dengan cara yang sistematis. Berikut adalah alur penelitian pada Gbr 1



Gbr 1. Proses Siklus KDD

A. Identifikasi Masalah

Pada tahap pertama, dilakukan identifikasi masalah dengan cara melakukan observasi pada *tweet* dari social media *Twitter* yang berkaitan dengan pinjaman online. Dari hasil observasi ditemukan permasalahan dari isu-isu di masyarakat mengenai pinjaman online yang terdapat pada *tweet*. Kurangnya pemahaman dari isu yang tersebar, sehingga diperlukan identifikasi dan pengelompokan topik yang sedang dibicarakan atau dibahas di sosial media untuk mengetahui dan memahami topik diskusi pinjaman online yang populer di *Twitter*

B. Studi Literatur

Tahap selanjutnya yaitu studi literatur untuk mendapatkan pemahaman tentang konsep, metode, dan teknologi yang relevan dengan masalah dan solusinya. Studi literatur juga dilakukan untuk mendapatkan informasi melalui literatur penelitian tentang masalah dan solusinya.

C. Knowledge Discovery in Databases (KDD)

Pada penelitian ini, digunakan metodologi *Knowledge Discovery in Databases* (KDD) untuk mengidentifikasi dan mengelompokkan isu-isu dari tweet mengenai pinjaman online. Menurut [12] Data mining adalah komponen penting dalam proses yang dikenal sebagai Knowledge Discovery in Databases (KDD) dan digunakan untuk mengidentifikasi dan mengekstrak informasi yang berharga, mudah dimengerti, dan baru dari kumpulan data yang besar dan rumit

1) Selection

Pada tahap ini dilakukan seleksi dengan cara melakukan pengumpulan, pemahaman, dan pemelihan data untuk menyiapkan dokumen yang akan dianalisis dengan menggunakan pemodelan topik.

2) Preprocessing

Dalam tahap preprocessing terdapat tahapan pengolahan data untuk melakukan pemodelan topik dengan metode LDA. Adapun pada tahapan *preprocessing* meliputi *remove user*, *cleaning data*, *case folding*, *tokenizing*, *normalisasi*, dan *stopword*.

3) Transformation

Tahap transformation, setelah tahap *preprocessing* teks selesai, langkah berikutnya adalah mengekstrak informasi penting dari data yang telah diproses. Tujuan dari proses ini adalah untuk meningkatkan efektivitas dan ketepatan pengolahan data.

4) Data Mining

Pada tahap ini dilakukan pemilihan teknik *data mining*. Setelah melakukan berbagai prosedur untuk mendapatkan komponen yang diperlukan, langkah selanjutnya adalah melanjutkan ke proses pelatihan model Latent Dirichlet Allocation (LDA).

5) Pattern Evaluation

Untuk memastikan bahwa pola atau model yang ditemukan selama tahap *data mining* akan dievaluasi untuk memastikan pola yang ditemukan valid. Pada tahap ini akan digunakan coherence score untuk mengevaluasi kualitas pemodelan topik.

6) Knowledge

Pada tahap terakhir, informasi yang ditemukan disampaikan dalam bentuk berupa visualisasi. Visualisasi yang digunakan yaitu visualisasi pyLDAvis dan *wordcloud* agar mudah diapahami. Ini menawarkan cara visual untuk memahami dan menafsirkan hasil pemodelan topik, menjadikannya alat yang berguna bagi para ilmuwan dan peneliti data pemrosesan bahasa alam (NLP)[13].

III. HASIL DAN PEMBAHASAN

Paragraf harus teratur. Semua paragraf harus rata, yaitu sama-sama rata kiri dan dan rata kanan.

A. Selection

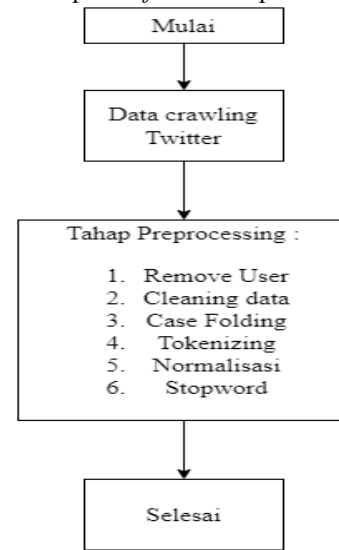
Pengambilan data menggunakan teknik *crawling* menggunakan Google Collab. Data yang digunakan adalah data *tweet* dari social media *Twitter* dengan kata kunci yang digunakan sebagai *keywords* yaitu “pinjol” dan “pinjaman online”, menggunakan Bahasa Indonesia mulai bulan Juli 2023 hingga Oktober 2023, didapatkan jumlah data 5172 data. Data yang didapat setelah proses penghapusan data duplikat dan data yang bernilai *null* terdapat 3435 data. Berikut proses penghapusan data duplikat dan bernilai *null* pada Gbr 2

	date tweets	username	id_tweets	Tweets
0	2023-08-01 03:52:43+00:00	gloutenc	1686223405778099968	yuk aku masih open pinpri yaa, bisa request no...
1	2023-08-01 05:45:50+00:00	debtcallcenter	1686251872666160128	@perfofm pinjaman pribadi kak, sama aja kek pi...
2	2023-08-01 06:27:42+00:00	AWPEPEK	1686262411618299904	@selebtwt Asli bang emang meresahkan, apalagi...
3	2023-08-01 10:47:11+00:00	strawberiff	1686327713102279936	cuy genuinely asking ki di endorse sama brand...
4	2023-08-01 13:21:53+00:00	iipsaepuloh80	1686366641150680064	@DivHumas_Polri Kalau pinjaman online itu term...

Gbr 2. Hasil data *Twitter*

B. Preprocessing

Adapun tahapan yang dilakukan pada *preprocessing*. Berikut *flowchart* tahap *transformation* pada Gbr 3



Gbr 3. Flowchart tahap *preprocessing*

1) Remove User

Pada tahap ini akan dilakukan untuk menghapus username *Twitter* (@username) dari teks *tweet*. Menggunakan fungsi ‘*remove_pattern*’ digunakan untuk menghilangkan pola teks tertentu dengan menggunakan ekspresi reguler. Dapat dilihat pada Tabel I

TABEL I
PERBANDINGAN REMOVE USER

Sebelum Remove User	Setelah Remove User
yuk aku masih open pinpri yaa, bisa request nominal japo lama langsung keteleku aja ya #pinpri #pinjamanpribadi #PinjamanOnline #zonauang	yuk aku masih open pinpri yaa, bisa request nominal japo lama langsung keteleku aja ya #pinpri #pinjamanpribadi #PinjamanOnline #zonauang
@perfofm pinjaman pribadi kak, sama aja kek pinjaman online tpi klo ini kita minjemnya ke org bukan aplikasi	pinjaman pribadi kak, sama aja kek pinjaman online tpi klo ini kita minjemnya ke org bukan aplikasi

2) Cleaning Data

Pada tahap ini akan dilakukan pembersihan data *cleaning data* untuk membersihkan teks dari *tweet*. Untuk melakukan identifikasi dan penghapusan pola-pola tertentu seperti, menghapus mention, menghapus hastag, menghapus indikator retweet, menghapus URL, menghapus angka, menghapus emoticon dan menghapus spasi ekstra. Dapat dilihat pada Tabel II

TABEL II
PERBANDINGAN CLEANING DATA

Sebelum <i>Cleaning Data</i>	Setelah <i>Cleaning Data</i>
yuk aku masih open pinpri yaa, bisa request nominal japo lama langsung keteleku aja ya #pinpri #pinjamanpribadi #PinjamanOnline #zonauang	yuk aku masih open pinpri yaa bisa request nominal japo lama langsung keteleku aja pinpri pinjamanpribadi PinjamanOnline zonauang
@perfowm pinjaman pribadi kak, sama aja kek pinjaman online tpi klo ini kita minjemnya ke org bukan aplikasi	pinjaman pribadi kak sama aja kek pinjaman online tpi klo ini kita minjemnya org bukan aplikasi

3) Case Folding

Tahap case folding digunakan untuk merubah karakter lurus kapital menjadi karakter huruf lain. Dalam pengolahan teks, proses ini sangat penting, terutama ketika kita ingin memperlakukan kata-kata yang sama, tanpa melihat apakah mereka ditulis dengan huruf kapital atau huruf kecil. Dapat dilihat pada Tabel III

TABEL III
PERBANDINGAN CASE FOLDING

Sebelum <i>Case Folding</i>	Setelah <i>Case Folding</i>
yuk aku masih open pinpri yaa bisa request nominal japo lama langsung keteleku aja pinpri pinjamanpribadi PinjamanOnline zonauang	yuk aku masih open pinpri yaa bisa request nominal japo lama langsung keteleku aja pinpri pinjamanpribadi pinjamanonline zonauang
pinjaman pribadi kak sama aja kek pinjaman online tpi klo ini kita minjemnya org bukan aplikasi	pinjaman pribadi kak sama aja kek pinjaman online tpi klo ini kita minjemnya org bukan aplikasi

4) Tokenizing

Pada tahap *tokenizing* dilakukan untuk memecah teks yang kompleks menjadi bagian yang lebih kecil, seperti kata, frasa, simbol, dan elemen lainnya, sehingga lebih mudah untuk melakukan analisis lebih lanjut. Dapat dilihat pada Tabel IV

TABEL IV
PERBANDINGAN TOKENIZING

Sebelum <i>Tokenizing</i>	Setelah <i>Tokenizing</i>
yuk aku masih open pinpri yaa	yuk aku masih open pinpri yaa

bisa request nominal japo lama langsung keteleku aja pinpri pinjamanpribadi PinjamanOnline zonauang	bisa request nominal japo lama langsung keteleku aja pinpri pinjamanpribadi pinjamanonline zonauang
pinjaman pribadi kak sama aja kek pinjaman online tpi klo ini kita minjemnya org bukan aplikasi	pinjaman pribadi kak sama aja kek pinjaman online tpi klo ini kita minjemnya org bukan aplikasi

5) Normalisasi

Untuk mengubah kalimat slang menjadi formal, menggunakan dan membaca dari kamus "colloquial-indonesian-lexicon.csv". Kamus ini mencakup kosa kata slang dan bentuk formal bahasa Indonesia. kata slang akan diganti dengan bentuk formalnya sesuai dengan kamus. Dapat dilihat pada Tabel V

TABEL V
PERBANDINGAN NORMALISASI

Sebelum Normalisasi	Setelah Normalisasi
yuk,aku,masih,open,pinpri,yaa,bisa,request,nominal,japo,lama,langsung,keteleku,aja,pinpri,pinjam anpribadi,pinjamanonline,zonauang	yuk,aku,masih,open,pinpri,ya,bisa,request,nominal,japo,lama,langsung,keteleku,saja,pinpri,pinjam anpribadi,pinjamanonline,zonauang
pinjaman,pribadi,kak,sama,ajak,ek,pinjeman,online,tpi,klo,ini,kita,minjemnya,ke,org,bukan,aplikasi	pinjaman,pribadi,kak,sama,saja,kayak,pinjeman,online,tapi,kalo,ini,kita,minjemnya,orang,bukan,aplikasi

6) Stopword

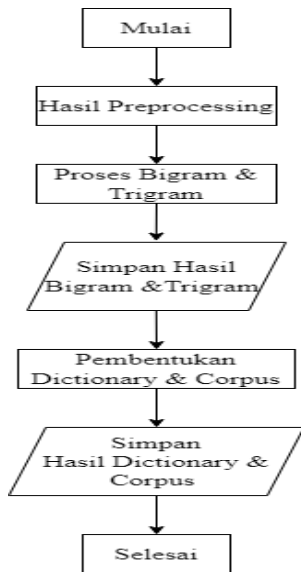
Pada tahap ini, *Stopword* dilakukan untuk mengeliminasi kata-kata yang dianggap tidak penting dalam teks. Dapat dilihat pada Tabel VI

TABEL VI
PERBANDINGAN STOPWORD

Sebelum Stopword	Setelah Stopword
yuk,aku,masih,open,pinpri,ya,bisa,request,nominal,japo,lama,langsung,keteleku,saja,pinpri,pinjamanpribadi,pinjamanonline,zonauang	pinpri,request,nominal,langsung,pinpri,pinjamanpribadi,pinjamanonline,zonauang
pinjaman,pribadi,kak,sama,saja,kayak,pinjeman,online,tapi,kalo,ini,kita,minjemnya,orang,bukan,aplikasi	pinjaman,pribadi,kak,kayak,pinjeman,online,minjemnya,orang,aplikasi

C. Transformation

Model bigram dan trigram adalah salah satu cara untuk mencapai hal ini. Kemudian pada tahap berikutnya akan dilakukan pembentukan vector representasi vector dari teks dalam dokumen dengan menggunakan *bag of words*. Berikut *flowchart* tahap *transformation* pada Gbr 4



Gbr 4. Flowchart tahap *transformation*

1) Bigram dan Trigram

Tujuan dari proses menggunakan model bigram dan trigram adalah untuk menemukan dan memahami makna yang terbentuk dari kombinasi dua atau tiga kata. Dalam analisis teks, pendekatan ini sangat penting karena makna sebuah ungkapan atau frasa seringkali hanya dapat dipahami secara keseluruhan ketika kata-kata tersebut dilihat bersama, bukan secara terpisah

TABEL VII
 HASIL BIGRAM DAN TRIGRAM

<i>Clean Tweets</i>	Bigram	Trigram
yuk aku masih open pinpri yaa bisa request nominal japo lama langsung keteleku aja ya	Pinpri,request,nominal,langsung,pinpri,pinjamanpribadi,pinjamanonline,zonauang	pinpri,request,nominal,langsung,pinpri,pinjamanpribadi,pinjamanonline,zonauang
pinjaman pribadi kak sama aja kek pinjaman online tpi klo ini kita minjemnya ke org bukan aplikasi	pinjaman_pribadi,kak,kayak,pinjeman,online,minjemnya,orang,aplikasi	pinjaman_pribadi,kak,kayak,pinjeman,online,minjemnya,orang,aplikasi

2) Pembentukan Dictionary dan Corpus

Sebelum memulai proses pemodelan topik menggunakan *Latent Dirichlet Allocation* (LDA), pembentukan corpus dan dictionary.

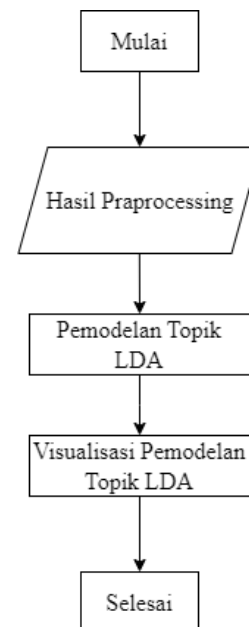
ID Kata	Kata	Frekuensi
0	langsung	2.0
1	nominal	2.0
2	pinjamanonline	2.0
3	pinjamanpribadi	2.0
4	pinpri	4.0
5	zonauang	2.0

Gbr 5. Hasil Dictionary dan Corpus

Dari Gbr 5 diketahui ID dan Kata yang merupakan hasil dari pembentukan *dictionary*. Sedangkan frekuensi merupakan hasil dari *corpus*. Hasil dari pembentukan *dictionary* dan *corpus* akan dimasukkan dalam pemodelan *Latent Dirichlet Allocation*.

D. Data Mining

Proses ini mencakup sejumlah langkah penting yang harus diikuti untuk memastikan bahwa model LDA dapat dilatih dengan baik dan efisien. Namun, sebelum melanjutkan pemodelan topik menggunakan algoritma LDA, terdapat tahapan untuk pembentukan dictionary dan corpus. Berikut tahapan yang dilakukan untuk pelatihan model LDA pada Gbr 6



Gbr 6. Flowchart tahap *data mining*

1) Implementasi Pemodelan *Latent Dirichlet Allocation* (LDA)

Tujuan dari tahap pembentukan topik model adalah untuk menghasilkan model topik terbaik untuk dokumen. Melakukan eksperimen pada nilai input parameter dilakukan untuk menghasilkan model topik yang tepat. Untuk

mendapatkan hasil pemodelan terbaik dan mudah diinterpretasi oleh manusia, maka dilakukan berbagai eksperimen pada *input* parameter. Pada tahapan eksperimen ini akan dilakukan skenario percobaan pada input nilai parameter yang berbeda setiap pemodelan yaitu *passes* atau jumlah itersai, nilai *alpha* dan *beta*. Untuk nilai parameter jumlah topik ditetapkan dengan nilai default *num of topic* adalah 10. Berikut nilai parameter pada Tabel VIII

TABEL VIII
 HASIL BIGRAM DAN TRIGRAM

Parameter	Nilai
Iterasi	10; 50; 100
Alpha	0,1; 0,001
Beta	0,1; 0,001

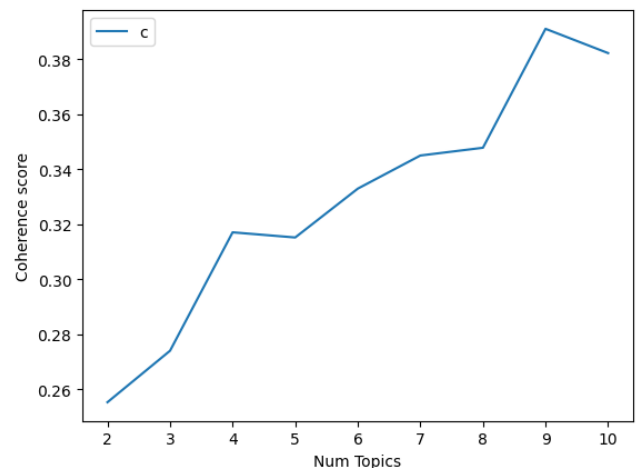
Hasil pemodelan topik akan dievaluasi menggunakan *coherence score* untuk membantu dalam menentukan jumlah topik yang optimal dari jumlah topik yang menghasilkan skor tertinggi dan paling masuk akal secara interpretative. Setelah dilakukan pertimbangan terhadap parameter dalam pemodelan topik dan perhitungan nilai tertinggi *coherence score* yang didapatkan dari hasil masing-masing pemodelan sebagai berikut pada Tabel IX

TABEL IX
 PERBANDINGAN EKSPERIMEN PARAMETER MODEL LDA

Skenario	Nilai Input Parameter			Coherence Score	Num Topics
	Iterasi	Alpha	Beta		
1	10	0,1	0,1	0.391	9
2	10	0,1	0,001	0,381	9
3	10	0,001	0,1	0.383	10
4	10	0,001	0,001	0.366	10
5	50	0,1	0,1	0.3713	10
6	50	0,1	0,001	0.376	9
7	50	0,001	0,1	0.382	10
8	50	0,001	0,001	0.366	9
9	100	0,1	0,1	0.376	9
10	100	0,1	0,001	0.379	9
11	100	0,001	0,1	0.377	10
12	100	0,001	0,001	0.368	9

Dari Tabel IX didapatkan hasil perbandingan dari skenario pemodelan topik dengan LDA yang didapatkan dari nilai *coherence score* tertinggi pada masing-masing skenario. Dimana dalam pemodelan LDA skenario 1 dengan input parameter untuk iterasi bernilai 10, nilai *alpha* sebesar 0,1 dan *beta* sebesar 0,1. Dalam pemodelan skenario 1 ini didapat nilai *coherence score* tertinggi sebesar 0,391. Nilai ini lebih unggul dari skenario lainnya dengan nilai tertinggi dan didapat jumlah topik yaitu 9 topik, dan input parameter pada skenario 1 akan menjadi acuan implementasi pemodelan LDA.

Untuk mendapatkan hasil pemodelan topik terbaik telah dilakukan berbagai eksperimen dengan setiap skenario memiliki parameter berbeda dan didapatkan pemodelan topik terbaik dengan mengetahui nilai *coherence score* tertinggi dan visualisasi grafik *coherence score*. *Coherence score* menjadi salah satu ukuran untuk mengevaluasi *topic modelling*, jika nilai yang didapat dari *coherence score* semakin tinggi maka model yang dibuat telah optimal. Setelah melakukan perbandingan dari eksperimen diatas dengan menentukan berbagai parameter yang terdiri dari nilai *K*, iterasi, *alpha* dan *beta*, didapatkan pemodelan topik terbaik yang terdapat pada skenario 1. Berikut hasil visualisasi *coherence score* yang ditunjukkan pada Gbr 7



Gbr 7 Grafik Coherence Score skenario 1

Setelah mendapatkan jumlah topik berdasarkan nilai dan grafik dari *coherence score* didapatkan model LDA pada Tabel X dengan banyaknya jumlah topik adalah 9 dengan penyebaran jumlah kata yang akan ditampilkan dalam pemodelan topik sebanyak 20 kata yang memiliki nilai bobot probabilitas setiap kata.

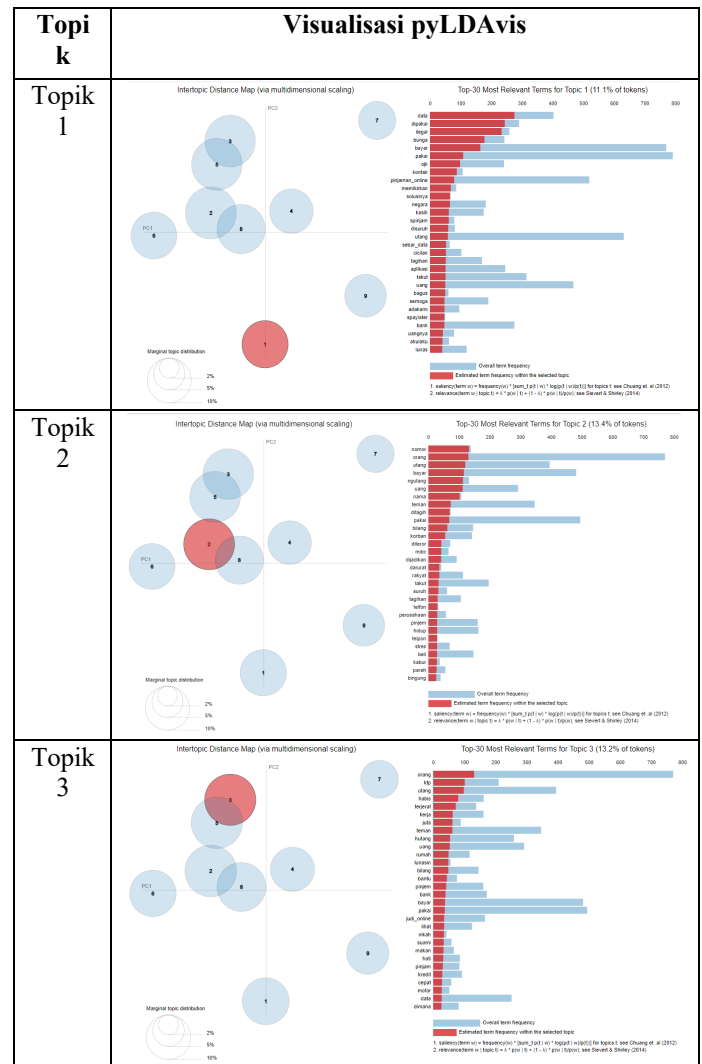
TABEL X
 HASIL LUARAN PEMODELAN MODEL LDA

<p>Model LDA Topik 1</p> <p>0.032*"data" + 0.028*"dipakai" + 0.027*"ilegal" + 0.020*"bunga" + 0.019*"bayar" + 0.012*"pakai" + 0.011*"ojk" + 0.010*"kontak" + 0.009*"pinjaman_online" + 0.008*"memikirkan" + 0.008*"solusinya" + 0.007*"negara" + 0.007*"kasih" + 0.007*"spinjam" + 0.007*"disuruh" + 0.007*"utang" + 0.006*"sebar_data" + 0.006*"cicilan" + 0.006*"tagihan" + 0.006*"aplikasi"</p>
<p>Model LDA Topik 2</p> <p>0.020*"nomor" + 0.020*"orang" + 0.018*"utang" + 0.018*"bayar" + 0.017*"ngutang" + 0.017*"uang" + 0.015*"nama" + 0.011*"teman" + 0.011*"ditagih" + 0.010*"pakai" + 0.009*"bilang" + 0.008*"korban" + 0.006*"diteror" + 0.006*"mikir" + 0.006*"dijadikan" + 0.005*"darurat" + 0.005*"rakyat" + 0.005*"takut" + 0.005*"suruh" + 0.005*"tagihan"</p>
<p>Model LDA Topik 3</p>

<p>0.020*"orang" + 0.016*"ktp" + 0.015*"utang" + 0.012*"habis" + 0.011*"terjerat" + 0.010*"kerja" + 0.010*"juta" + 0.009*"teman" + 0.008*"hutang" + 0.008*"uang" + 0.008*"rumah" + 0.008*"lunasin" + 0.007*"bilang" + 0.007*"bantu" + 0.006*"pinjem" + 0.006*"bank" + 0.006*"bayar" + 0.006*"pakai" + 0.005*"judi_online" + 0.005*"lihat"</p> <p>Model LDA Topik 4</p>
<p>0.012*"legal" + 0.012*"tutup" + 0.011*"judi" + 0.011*"ojk" + 0.010*"aplikasi" + 0.010*"teman" + 0.010*"bayar" + 0.009*"anak" + 0.009*"guru" + 0.008*"takut" + 0.008*"semoga" + 0.007*"lobang" + 0.007*"lubang" + 0.007*"nomer" + 0.007*"ganti" + 0.006*"gaji" + 0.006*"gali" + 0.006*"minjem" + 0.006*"indonesia" + 0.006*"twitter"</p> <p>Model LDA Topik 5</p>
<p>0.047*"orang" + 0.014*"kelilit" + 0.011*"pakai" + 0.011*"bayar" + 0.010*"susah" + 0.009*"salah" + 0.009*"bunuh" + 0.009*"hidup" + 0.008*"anak" + 0.007*"teman" + 0.007*"dimana" + 0.006*"lihat" + 0.006*"hati" + 0.006*"judi_online" + 0.006*"miskin" + 0.006*"habis" + 0.006*"main" + 0.005*"usaha" + 0.005*"gila" + 0.005*"rakyat"</p> <p>Model LDA Topik 6</p>
<p>0.039*"galbay" + 0.018*"cari" + 0.014*"pinpri" + 0.014*"bayar" + 0.013*"info" + 0.011*"buka" + 0.010*"telbay" + 0.009*"zonauang" + 0.009*"butuh" + 0.009*"teman" + 0.009*"pinjem" + 0.009*"daftar" + 0.009*"pinjaman" + 0.008*"orang" + 0.006*"takut" + 0.006*"susah" + 0.006*"mengobrol" + 0.006*"jalan" + 0.006*"utang" + 0.006*"uang"</p> <p>Model LDA Topik 7</p>
<p>0.022*"main" + 0.015*"utang" + 0.013*"capek" + 0.011*"beli" + 0.010*"kerja" + 0.009*"ketemu" + 0.009*"judi" + 0.008*"tukang" + 0.008*"kalah" + 0.008*"pakai" + 0.007*"rela" + 0.007*"grup" + 0.007*"bawa" + 0.007*"korban" + 0.007*"terjerat" + 0.006*"jual" + 0.006*"hasil" + 0.006*"dibayar" + 0.005*"verifikasi" + 0.005*"bundir"</p> <p>Model LDA Topik 8</p>
<p>0.030*"pakai" + 0.019*"minjem" + 0.018*"akun" + 0.015*"pinjaman_online" + 0.015*"orang" + 0.013*"jaminan" + 0.013*"hutang" + 0.012*"ktp" + 0.012*"coba" + 0.011*"riba" + 0.011*"online" + 0.010*"jadikan" + 0.009*"pinjaman" + 0.009*"teman" + 0.009*"rentenir" + 0.007*"dikejar" + 0.006*"tolong" + 0.006*"bunga" + 0.006*"bahaya" + 0.006*"seram"</p> <p>Model LDA Topik 9</p>
<p>0.043*"iklan" + 0.024*"pinjaman_online" + 0.015*"kemarin" + 0.013*"hutang" + 0.010*"dijadikan" + 0.008*"kasih" + 0.007*"paylater" + 0.007*"orang" + 0.007*"beda" + 0.006*"indonesia" + 0.006*"daftarin" + 0.006*"main" + 0.006*"minggu" + 0.006*"pakai" + 0.006*"korban" + 0.005*"langsung" + 0.005*"masyarakat" + 0.005*"aplikasi" + 0.005*"foto" + 0.005*"iklannya"</p>

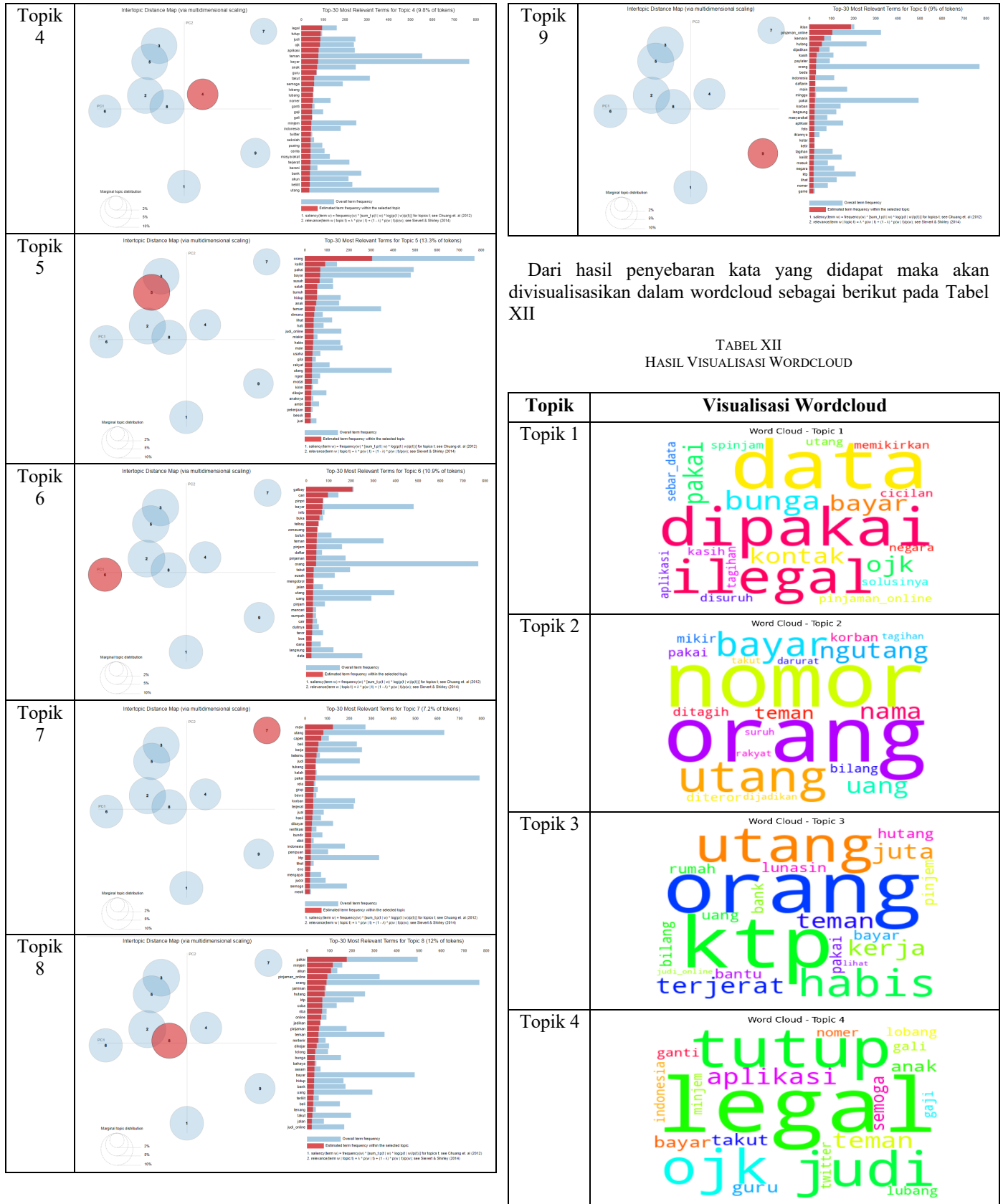
adalah *intertopic distance map* atau pemetaan jarak dari antar topik dengan bentuk lingkaran dan terdapat nomor yang merupakan topik. Dan pada panel sisi kanan menunjukkan *bar chart top 30 most salient term* atau 30 term yang paling relevan dengan topik tersebut. Untuk visualisasi wordcloud akan ditampilkan 20 kata teratas dari topik tersebut. Berikut hasil visualisasi dan interpretasi dari setiap topik pada Tabel X1

TABEL X1
HASIL VISUALISASI PYLDAVIS



2) Visualisasi dan Interpretasi Topik

Setelah dibuat pemodelan topik akan dibuat visualisasi dengan pyLDAvis dan wordcloud, yang nantinya hasil dari visualisasi tersebut akan diinterpretasikan setiap topik dari sebaran kata yang dihasilkan. Visualisasi pyLDAvis ini terdiri dari paneli sisi kiri dan panel sisi kanan, dimana panel sisi kiri



Dari hasil penyebaran kata yang didapat maka akan divisualisasikan dalam wordcloud sebagai berikut pada Tabel XII

TABEL XII
HASIL VISUALISASI WORDCLOUD

Topik	Visualisasi Wordcloud
Topik 1	Word Cloud - Topic 1 Data, bunga, bayar, dipakai, ilegal, pinjaman, utang, aplikasi, tagihan, kontak, negara, disuruh, pinjaman_online, kasihan, lojk, solusinya
Topik 2	Word Cloud - Topic 2 Nomor, orang, utang, bayaran, korban, tagihan, mikir, pakai, takut, darurat, ditagih, teman, nama, suruh, rakyat, utang, bilang, diteror, dijadikan, uang
Topik 3	Word Cloud - Topic 3 Utang, orang, hutang, rumah, lunasin, jutaan, bank, teman, ktp, bayar, kerja, judi_online, bantu, lihat, terjerat, habis, bilang
Topik 4	Word Cloud - Topic 4 Tutup, legal, aplikasi, minjem, bayartakut, ojk, judi, teman, semoga, gaj, lubang, ganti, indonesi, minjem, aplikasi, bayartakut, ojk, judi, teman, semoga, gaj, lubang



TABEL XIII
HASIL ISU TOPIK

Topik	Isu Topik
Topik 1	Data pengguna pinjaman online
Topik 2	Teror nomor pengaih pinjaman online
Topik 3	Penggunaan KTP untuk pengajuan injaman online
Topik 4	Aplikasi pinjaman online
Topik 5	Motif pengajuan pinjaman online
Topik 6	Kesulitan membayar pinjaman online
Topik 7	Korban judi online dan pinjaman online
Topik 8	Akun pengguna pinjaman online
Topik 9	Iklan pinjaman online

IV. KESIMPULAN

Penelitian ini bertujuan untuk mengetahui membantu dalam mengidentifikasi dan mengelompokan isu-isu dari *tweet* mengenai pinjaman online dengan metode *Latent Dirichlet Allocation* (LDA), didapat kesimpulan sebagai berikut:

1. Implementasi metode *Latent Dirichlet Allocation* (LDA) untuk membantu dalam mengidentifikasi dan mengelompokan isu-isu dari *tweet* mengenai pinjaman online memiliki data sebanyak 3435 *tweet*. Dari 12 kali eksperimen didapatkan pemodelan terbaik pada skenario 1. Dimana dalam pemodelan LDA skenario 1 paramater untuk iterasi bernilai 10, nilai alpha sebesar 0,1 dan beta sebesar 0,1. Dalam pemodelan scenario 2 ini didapat nilai coherence score tertinggi sebesar 0,391. Nilai ini lebih unggul dari sekenario lainnya dengan nilai tertinggi dan didapat jumlah topik yaitu 9 topik.
2. Hasil visualisasi dan interpretasi yang didapat dari visualisasi dalam *pyLDAvis* dan *wordcloud* didapatkan penyebaran isu yang berbeda setiap topik. Topik 1 mengenai data pengguna pinjaman online. Topik 2 membahas Teror nomor pengaih pinjaman online. Topik 3 membahas penggunaan KTP untuk pengajuan pinjaman online. Topik 4 mengenai aplikasi pinjaman online. Topik 5 membahas motif pengajuan pinjaman online. Topik 6 mengenai kesulitan membayar pinjaman online. Topik 7 membahas korban judi online dan pinjaman online. Topik 8 mengenai akun pengguna pinjaman online. Topik 9 membahas iklan pinjaman online

V. SARAN

Terdapat saran yang dapat digunakan untuk melakukan pengembangan pada penelitian selanjutnya yaitu:

1. Pada penelitian selanjutnya, diharapkan dapat

Dari visualisasi *pyLDAvis* dan *wordcloud*, didapatkan hasil interpretasi pada setiap topik. Hal ini dapat dilihat dari penyebaran kata yang memiliki bobot nilai tertinggi pada model tersebut. Berikut pada Tabel XIII hasil isu setiap topik.

menggunakan data dari platform lainnya seperti Facebook, Youtube, Instagram, atau TikTok.

2. Dari hasil pengamatan pada tahap preprocessing diperlukan adanya penyempurnaan terutama pada tahap stopword karena banyak kata-kata slank untuk itu perlu diperhatikan lagi untuk menghilangkan kata yang tidak memiliki makna. Pada penelitian selanjutnya dapat menambahkan tahap stemming pada preprocessing.

VI. UCAPAN TERIMA KASIH

Penulis menyampaikan apresiasi yang mendalam kepada semua pihak yang telah mendukung dan membantu dalam menyelesaikan penelitian ini. Terima kasih kepada Ibu Wiyli Yustanti atas bimbingan dan nasihat berharga yang diberikan sepanjang proses penelitian. Penghargaan juga ditujukan kepada kampus tercinta Universitas Negeri Surabaya yang telah menyediakan fasilitas dan dukungan finansial untuk penelitian ini. Tak lupa, penulis mengucapkan terima kasih kepada keluarga dan teman-teman atas dukungan moral dan semangat yang diberikan.

REFERENSI

- [1] H. Susanto, Leu Fang Yie, Fadzliwati Mohiddin, Arief Amier Rahman Setiawan, Parastou Khodaparast Haghi, and Desi Setiana, "Revealing Social Media Phenomenon in Time of COVID-19 Pandemic for Boosting Start-Up Businesses through Digital Ecosystem," *Appl. Syst. Innov.*, Jan. 2021.
- [2] Asmah Savitri, Angga Syahputra, Husna Hayati, and Heny Rofizar, "Pinjaman Online di Masa Pandemi Covid-19 bagi Masyarakat Aceh," *E-Mabis J. Ekon. Manaj. dan Bisnis*, vol. 22 (2), Oct. 2021.
- [3] M Rahman Adinata and Recca Ayu Hapsar, "Tinjauan Yuridis Perlindungan Konsumen Terhadap Masyarakat Yang Melakukan Fintech Peer-To-Peer Lending Atau Layanan Pinjam Meminjam Uang Berbasis Teknologi Informasi (Lpmubti) Dalam Sebuah Aplikasi Pinjaman Online (Julo) (Studi Penelitian : Otoritas Jasa)," *Case law-Journal law*, vol. 3, Jan. 2022.
- [4] C. M. Annur, "Ada 3,9 Ribu Aduan Kasus Pinjol Ilegal sejak Awal 2023, Ini Tren Bulanannya," databoks. Accessed: Jan. 02, 2024. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2023/06/14/ada-39-ribu-aduan-kasus-pinjol-ilegal-sejak-awal-2023-ini-tren-bulanannya>
- [5] A. C. Mutia, "Jumlah Pengguna Twitter di Indonesia Capai 14,75 Juta per April 2023, Peringkat Keenam Dunia," databoks. Accessed: Nov. 01, 2023. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2023/05/31/jumlah-pengguna-twitter-di-indonesia-capai-1475-juta-per-april-2023-peringkat-keenam-dunia#:~:text=Media-,Jumlah Pengguna Twitter di Indonesia Capai 14%2C75 Juta,April 2023%2C Peringkat Keenam Dunia&t>
- [6] A. Alamsyah, W. Rizkika, D. D. A. Nugroho, F. Renaldi, and S. Saadah, "Dynamic Large Scale Data on Twitter using Sentiment Analysis and Topic Modeling: Uber," *2018 6th Int. Conf. Inf. Commun. Technol.*, vol. 0, pp. 254–258, 2018.
- [7] G. Nair, "Text mining 101: Topic modeling," *kdnuggets*. Accessed: Nov. 01, 2023. [Online]. Available: <http://www.kdnuggets.com/2016/07/text-mining-101-topicmodeling.html>
- [8] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet Allocation," *J. Mach. Learn. Res.* 3, pp. 993–1022, 2003.
- [9] H. Jelodar *et al.*, "Latent Dirichlet Allocation (LDA) and Topic modeling: models, applications, a survey," 2018.
- [10] F. Rashif, G. I. P. Nirvana, M. A. Noor, and N. A. Rakhmawati, "Implementasi LDA untuk Pengelompokan Topik Cuitan Akun Bot Twitter bertagar #Covid-19," *Cogito Smart Journa*, vol. 7, no. 1, 2021.
- [11] M. L. C. Chilmi, "Latent Dirichlet Allocation (LDA) Untuk mengetahui Topik Pembicaraan Warganet Tentang Omnibus Law," Universitas Islam Negeri Syarif Hidayatullah., Jakarta, 2021.
- [12] S. Widaningsih, "Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4.5, Naïve Bayes, Knn, Dan Svm," *J. Tekno Insentif*, vol. 13, 2019.
- [13] B. Mebey, "pyLDAvis," Sphinx. Accessed: Jan. 02, 2024. [Online]. Available: <https://pyldavis.readthedocs.io/en/latest/readme.html>