

**IMPLEMENTASI *TOPIC MODELING* DENGAN
METODE *LATENT DIRICHLET ALLOCATION* (LDA)
UNTUK MENGIDENTIFIKASI KARAKTERISTIK
KASUS TINDAK PIDANA PENCURIAN**

(Studi Kasus : Dokumen Surat Dakwaan Tindak Pidana Pencurian di Pengadilan
Negeri Palembang Tahun 2019 - 2022)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Program
Studi Statistika



Disusun Oleh:

Rizqia Azzahra Suci Ramadhanti

19611086

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2023**

**HALAMAN PERSETUJUAN PEMBIMBING
TUGAS AKHIR**

Judul : Implementasi *Topic Modeling* Dengan Metode *Latent Dirichlet Allocatin* (LDA) untuk Mengidentifikasi Karakteristik Kasus Tindak Pidana Pencurian (Studi Kasus : Dokumen Surat Dakwaan Tindak Pidana Pencurian di Pengadilan Negeri Palembang Tahun 2019 – 2022).

Nama Mahasiswa : Rizqia Azzahra Suci Ramadhanti

NIM : 19611086

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

Yogyakarta, 07 Maret 2023

Pembimbing



(Ayundyah Kesumawati, S.Si., M.Si.)

HALAMAN PENGESAHAN
TUGAS AKHIR

IMPLEMENTASI *TOPIC MODELING* DENGAN METODE *LATENT DIRICHLET ALLOCATION* (LDA) UNTUK MENGIDENTIFIKASI KARAKTERISTIK KASUS TINDAK PIDANA PENCURIAN
(Studi Kasus : Dokumen Surat Dakwaan Tindak Pidana Pencurian di Pengadilan Negeri Palembang Tahun 2019-2022)

Nama Mahasiswa : Rizqia Azzahra Suci Ramadhanti
NIM : 19611086

**TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL : 06 April 2023**

Nama Penguji

Tanda Tangan

1. Muhammad Hasan Sidiq Kurniawan, S.Si.,
M.Sc.

2. Dina Tri Utari, S.Si., M.Sc.

3. Ayundyah Kesumawati, S.Si., M.Si.

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



(Prof. Riyanto, S.Pd., M.Si., Ph.D.)



KATA PENGANTAR



Assalamu'alaikum Wr. Wb

Segala puji dan syukur bagi Allah SWT, atas segala rahmat dan karunia-Nya sehingga peneliti dapat menyelesaikan Tugas Akhir yang berjudul “Implementasi *Topic Modeling* Dengan Metode *Latent Dirichlet Allocatin* (LDA) untuk Mengidentifikasi Karakteristik Kasus Tindak Pidana Pencurian (Studi Kasus : Dokumen Surat Dakwaan Tindak Pidana Pencurian di Pengadilan Negeri Palembang Tahun 2019 – 2022)” sebagai syarat bagi mahasiswa untuk memperoleh gelar Sarjana Statistika pada jenjang Strata Satu (S1) Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.

Dalam proses penyusunan Tugas Akhir ini, peneliti menerima banyak sekali bimbingan, bantuan, dan dukungan dari banyak pihak sehingga Tugas Akhir ini dapat terselesaikan. Oleh karena itu, pada kesempatan ini peneliti ingin mengucapkan terima kasih kepada:

1. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D. Selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
2. Bapak Dr. Edy Widodo, S.Si., M.Si. Selaku Ketua Jurusan Statistika Universitas Islam Indonesia.
3. Ibu Dr. Atina Ahdika, S.Si., M.Si. Selaku Ketua Program Studi Statistika Universitas Islam Indonesia.
4. Ibu Ayundyah Kesumawati, S.Si., M.Si. Selaku Dosen Pembimbing Tugas Akhir yang telah meluangkan waktu untuk memberikan bimbingan, juga memberikan ilmu, arahan, motivasi serta saran untuk peneliti.
5. Kepada kedua orang tua, serta saudara-saudari peneliti, yang telah memberikan saran dan dukungan dalam penyusunan Tugas Akhir ini.
6. Bapak dan Ibu Dosen Program Studi Statistika, Universitas Islam Indonesia yang telah memberikan ilmu dan wawasannya kepada peneliti.
7. Teman-teman seperjuangan yaitu Monica Rahma Fauzia, Mitha Rahma Fadila, Alfa Yuliana, Rahmi Anadra, Dela Gustiara, dan Nadilah Fadia Haya, yang telah berjuang bersama dalam penyusunan Tugas Akhir.

8. Semua pihak terkait yang tidak bisa penulis sebutkan satu per satu, yang telah membantu secara langsung, maupun tidak langsung dalam penyusunan Tugas Akhir ini.

Akhir kata peneliti mengharapkan semoga Tugas Akhir ini dapat berguna untuk menambah wawasan bagi berbagai pihak, peneliti menyadari Tugas Akhir ini masih jauh dari kata sempurna, sehingga peneliti mengharapkan kritik dan saran yang membangun dalam penelitian ini.

Wassalamualaikum Wr.Wb

Yogyakarta, 06 April 2023



Rizqia Azzahra Suci Ramadhanti

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN PEMBIMBING TUGAS AKHIR.....	ii
HALAMAN PENGESAHAN TUGAS AKHIR	iii
KATA PENGANTAR.....	iv
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR.....	x
DAFTAR LAMPIRAN	xii
PERNYATAAN	xiii
INTISARI.....	xiv
ABSTRACT	xv
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	4
1.4. Tujuan Penelitian	4
1.5. Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA.....	6
BAB III LANDASAN TEORI.....	11
3.1. Tindak Pidana.....	11
3.2. Pencurian.....	11
3.2.1 Pencurian Biasa	12
3.2.2 Pencurian Ringan.....	12
3.2.3 Pencurian dengan Pemberatan.....	13
3.2.4 Pencurian dengan Kekerasan.....	13
3.2.5 Pencurian Dalam Kalangan Keluarga	13
3.3. Surat Dakwaan	13
3.4. Statistika Deskriptif.....	13
3.5. <i>Data Mining</i>	14
3.6. <i>Text Mining</i>	14
3.7. <i>Preprocessing</i>	15
3.8. <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i>	16
3.9. <i>Topic Modeling</i>	18
3.10. <i>Latent Dirichlet Allocation (LDA)</i>	18
3.11. <i>Topic Coherence</i>	22
3.12. <i>Wordcloud</i>	23
3.13. <i>Principal Component Analysis</i>	23
BAB IV METODOLOGI PENELITIAN	25
4.1. Populasi Penelitian	25
4.2. Jenis dan Sumber Data	25
4.3. Variabel penelitian	25
4.4. Metode dan Analisis Data	26
4.5. Tahapan Penelitian	26
BAB V HASIL DAN PEMBAHASAN.....	27
5.1. Analisis Deskriptif	27

5.2.	<i>Preprocessing Data</i>	29
5.2.1	<i>Case Folding</i>	31
5.2.2	<i>Remove Punctuation</i>	34
5.2.3	<i>Remove Number</i>	36
5.2.4	<i>Stopword Removal</i>	39
5.2.5	<i>Stemming</i>	41
5.2.6	<i>Tokenizing</i>	42
5.3.	Pembobotan Kata TF-IDF.....	42
5.4.	<i>Topic Modeling dengan Latent Dirichlet Allocation</i>	45
5.4.1	Tahun 2019.....	45
5.4.2	Tahun 2020.....	52
5.4.3	Tahun 2021.....	64
5.4.4	Tahun 2022.....	73
5.5.	Perbandingan Topik	83
	BAB VI PENUTUP	86
6.1.	Kesimpulan	86
6.2.	Saran.....	87
	DAFTAR PUSTAKA.....	88
	LAMPIRAN	93

DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Sebelumnya	8
Tabel 4.1 Variabel Penelitian	25
Tabel 5.1 Data Awal.....	30
Tabel 5.2 <i>Case Folding</i>	31
Tabel 5.3 <i>Remove Punctuation</i>	34
Tabel 5.4 <i>Remove Number</i>	36
Tabel 5.5 <i>Stopword Removal</i>	39
Tabel 5.6 <i>Stemming</i>	41
Tabel 5.7 <i>Tokenizing</i>	42
Tabel 5.8 Sampel Hasil Pembobotan Kata TF-IDF Tahun 2022	43
Tabel 5.9 Hasil Perhitungan Nilai TF	44
Tabel 5.10 Hasil Pembobotan Kata Mobil	45
Tabel 5.11 Hasil Nilai <i>Coherence</i> Tahun 2019.....	46
Tabel 5.12 Hasil Nilai PC Tahun 2019	47
Tabel 5.13 Model Topik 1 Tahun 2019.....	49
Tabel 5.14 Model Topik 2 Tahun 2019.....	50
Tabel 5.15 Model Topik 3 Tahun 2019.....	51
Tabel 5.16 Hasil Nilai <i>Coherence</i> Tahun 2020.....	53
Tabel 5.17 Hasil Nilai PC Tahun 2020	54
Tabel 5.18 Model Topik 1 Tahun 2020.....	56
Tabel 5.19 Model Topik 2 Tahun 2020.....	57
Tabel 5.20 Model Topik 3 Tahun 2020.....	58
Tabel 5.21 Model Topik 4 Tahun 2020.....	60
Tabel 5.22 Model Topik 5 Tahun 2020.....	61
Tabel 5.23 Model Topik 6 Tahun 2020.....	62
Tabel 5.24 Hasil Nilai <i>Coherence</i> Tahun 2021	64
Tabel 5.25 Hasil Nilai PC Tahun 2021	65
Tabel 5.26 Model Topik 1 Tahun 2021.....	67
Tabel 5.27 Model Topik 2 Tahun 2021.....	68
Tabel 5.28 Model Topik 3 Tahun 2021.....	70

Tabel 5.29 Model Topik 4 Tahun 2021.....	71
Tabel 5.30 Model Topik 5 Tahun 2021.....	72
Tabel 5.31 Hasil Nilai <i>Coherence</i> Tahun 2022.....	74
Tabel 5.32 Hasil Nilai PC Tahun 2022	75
Tabel 5.33 Model Topik 1 Tahun 2022.....	77
Tabel 5.34 Model Topik 2 Tahun 2022.....	78
Tabel 5.35 Model Topik 3 Tahun 2022.....	79
Tabel 5.36 Model Topik 4 Tahun 2022.....	81
Tabel 5.37 Model Topik 5 Tahun 2022.....	82
Tabel 5.38 Kesimpulan Topik Per Tahun	84
Tabel 5.39 Hasil Perbandingan Topik Antar Tahun	84

DAFTAR GAMBAR

Gambar 1.1 Data Jenis Kejahatan Tertinggi di Indonesia Periode Tahun 2019-2022 (Sumber: Pusiknas Bareskrim Polri).....	1
Gambar 3.1 <i>Plate Notation</i> LDA (Sumber: Setijohatmo et al., 2020)	19
Gambar 4.1 Diagram Alir Penelitian.....	26
Gambar 5.1 Waktu Terjadinya Pencurian Tahun 2019 – Tahun 2022.....	27
Gambar 5.2 Tindak Pidana yang Ditetapkan Kepada Terdakwa	28
Gambar 5.3 Rata-Rata Hukuman (Bulan) Kasus Pencurian Tahun 2019 – Tahun 2022.....	29
Gambar 5.4 Grafik Nilai <i>Coherence</i> Tahun 2019	46
Gambar 5.5 Visualisasi Ketiga Topik Tahun 2019	48
Gambar 5.6 Visualisasi Topik 1 Tahun 2019.....	49
Gambar 5.7 Visualisasi Topik 2 Tahun 2019.....	50
Gambar 5.8 Visualisasi Topik 3 Tahun 2019.....	51
Gambar 5.9 <i>Wordcloud</i> Topik 2.....	52
Gambar 5.10 Grafik Nilai <i>Coherence</i> Tahun 2020	53
Gambar 5.11 Visualisasi Keenam Topik Tahun 2020.....	55
Gambar 5.12 Visualisasi Topik 1 Tahun 2020.....	56
Gambar 5.13 Visualisasi Topik 2 Tahun 2020.....	57
Gambar 5.14 Visualisasi Topik 3 Tahun 2020.....	58
Gambar 5.15 Visualisasi Topik 4 Tahun 2020.....	59
Gambar 5.16 Visualisasi Topik 5 Tahun 2020.....	61
Gambar 5.17 Visualisasi Topik 6 Tahun 2020.....	62
Gambar 5.18 <i>Wordcloud</i> Topik 2.....	63
Gambar 5.19 Grafik Nilai <i>Coherence</i> Tahun 2021	64
Gambar 5.20 Visualisasi Kelima Topik Tahun 2021	66
Gambar 5.21 Visualisasi Topik 1 Tahun 2021	67
Gambar 5.22 Visualisasi Topik 2 Tahun 2021	68
Gambar 5.23 Visualisasi Topik 3 Tahun 2021	69
Gambar 5.24 Visualisasi Topik 4 Tahun 2021	70
Gambar 5.25 Visualisasi Topik 5 Tahun 2021	72

Gambar 5.26 <i>Wordcloud</i> Topik 5.....	73
Gambar 5.27 Grafik Nilai <i>Coherence</i> Tahun 2022	74
Gambar 5.28 Visualisasi Kelima Topik Tahun 2022	76
Gambar 5.29 Visualisasi Topik 1 Tahun 2022.....	77
Gambar 5.30 Visualisasi Topik 2 Tahun 2022.....	78
Gambar 5.31 Visualisasi Topik 3 Tahun 2022.....	79
Gambar 5.32 Visualisasi Topik 4 Tahun 2022.....	80
Gambar 5.33 Visualisasi Topik 5 Tahun 2022.....	82
Gambar 5.34 <i>Wordcloud</i> Topik 4.....	83

DAFTAR LAMPIRAN

Lampiran 1	93
Lampiran 2	93

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 06 April 2023



Rizqia Azzahra Suci Ramadhanti

INTISARI

IMPLEMENTASI *TOPIC MODELING* DENGAN METODE *LATENT DIRICHLET ALLOCATION (LDA)* UNTUK MENGIDENTIFIKASI KARAKTERISTIK KASUS TINDAK PIDANA PENCURIAN

(Studi Kasus : Dokumen Surat Dakwaan Tindak Pidana Pencurian di Pengadilan Negeri Palembang Tahun 2019-2022)

Rizqia Azzahra Suci Ramadhanti

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

Pencurian adalah perbuatan mengambil benda atau harta milik orang lain tanpa seizin pemilikinya, tindakan pencurian merupakan salah satu tindak pidana yang sangat meresahkan masyarakat. Menurut Kapolrestabes Palembang pada tahun 2020, dari berbagai macam kejahatan, terdapat 4 kejahatan yang penyelesaian kasusnya menjadi prioritas dikarenakan kejahatan tersebut membuat resah masyarakat, yaitu pencurian dengan pemberatan, pencurian dengan kekerasan, pencurian kendaraan bermotor, dan penganiayaan berat. Agar dapat membantu masyarakat dalam memahami cara kerja para pelaku pencurian serta dapat mencegah tindak pidana pencurian di masa yang akan datang adalah dengan mengetahui karakteristik dari kasus pencurian yang telah terjadi. Dengan diketahuinya karakteristik pencurian, masyarakat dapat memahami bagaimana seorang pencuri beraksi, sehingga dapat lebih waspada dan mengambil tindak pencegahan agar menghindari dan meminimalisir risiko menjadi korban pencurian. Oleh karena itu, peneliti melakukan analisis *Topic Modeling* dengan menggunakan metode *Latent Dirichlet Allocation (LDA)* untuk mengidentifikasi karakteristik kasus tindak pidana pencurian pada dokumen surat dakwaan kasus tindak pidana pencurian di Pengadilan Negeri Palembang tahun 2019-2022. Diperoleh jumlah topik yang terbentuk pada tahun 2019, 2020, 2021, dan 2022 secara berturut-turut adalah 4 topik, 5 topik, 6 topik, dan 4 topik. Berdasarkan perbandingan topik antar tahun, didapatkan pada tahun 2020 karakteristik baru yang didapatkan berupa korban yang menjadi sasaran pencurian dan tempat terjadi pencurian, pada tahun 2021 karakteristik baru yang didapatkan berupa jenis pencurian, yaitu pencurian dengan kekerasan dan pencurian dengan bersekutu, dan tidak terdapat perbedaan karakteristik yang didapatkan pada tahun 2022 dengan tahun sebelumnya terkait karakteristik kasus tindak pidana pencurian.

Kata Kunci : *Topic Modeling, Latent Dirichlet Allocation, Tindak Pidana Pencurian.*

ABSTRACT

IMPLEMENTATION OF TOPIC MODELING WITH LATENT DIRICHLET ALLOCATION (LDA) METHOD TO IDENTIFY CHARACTERISTICS OF THEFT CRIME CASES

(Case Study : Document of Indictment of Theft Crime Cases in Palembang District Court 2019-2022)

Rizqia Azzahra Suci Ramadhanti
Department of Statistics, Faculty of Matematics and Natural Sciences
Universitas Islam Indonesia

Theft is the act of taking other people's objects or property without the permission of the owner, the act of theft is one of the criminal acts that is very troubling to the community. According to the Palembang Police Chief in 2020, of the various crimes, there are 4 crimes whose case resolution is a priority because these crimes make the community restless, namely burglary, robbery, motor vehicle theft, and severe persecution. To help the community understand how the perpetrators of theft perpetrators and can prevent theft crimes in the future is to know the characteristics of theft cases that have occurred. By knowing the characteristics of theft, the public can understand how a thief acts, so that they can be more vigilant and take precautions to avoid and minimize the risk of becoming a victim of theft. Therefore, researchers conducted a Topic Modeling analysis using the Latent Dirichlet Allocation (LDA) method to identify the characteristics of theft criminal cases in the indictment documents of theft cases in the Palembang District Court in 2019-2022. The number of topics formed in 2019, 2020, 2021, and 2022 respectively is 4 topics, 5 topics, 6 topics, and 4 topics. Based on the comparison of topics between years, it was obtained in 2020 that the new characteristics obtained were in the form of victims who were targeted by theft and where theft occurred, in 2021 the new characteristics obtained were in the form of types of theft, namely theft with and theft with allies, and there was no difference in the characteristics obtained in 2022 with the previous year related to the characteristics of criminal theft cases.

Keywords: *Topic Modeling, Latent Dirichlet Allocation, Theft Crime.*

BAB I

PENDAHULUAN

1.1. Latar Belakang

Kejahatan merupakan suatu tindakan atau perbuatan yang dilarang dan dapat dijatuhi hukuman atas-nya oleh Undang-Undang. Pandangan itu dilihat dari suatu teori yang menyatakan bahwa setiap anggota masyarakat adalah makhluk yang mempunyai kehendak bebas. Menurut Topo Santoso dalam jurnal (Arif, 2014), secara sosiologis, kejahatan merupakan suatu kepribadian manusia yang dibentuk oleh masyarakat, walaupun masyarakat memiliki berbagai macam perilaku yang berbeda-beda akan tetapi ada didalamnya bagian tertentu yang mempunyai pola yang sama.

Berdasarkan data Pusat Informasi Kriminal Nasional (Pusiknas) Kepolisian Indonesia (Polri) terdapat 962.096 kasus kejahatan yang terjadi sepanjang tahun 2019 hingga 2022, dimana pada tahun 2019 sebanyak 178.837 kasus, tahun 2020 sebanyak 188.648, tahun 2021 sebanyak 275.164, dan tahun 2022 sebanyak 319.447 kasus kejahatan yang terjadi di Indonesia.



Gambar 1.1 Data Jenis Kejahatan Tertinggi di Indonesia Periode Tahun 2019-2022 (Sumber: Pusiknas Bareskrim Polri)

Gambar 1.1 menunjukkan bahwa jenis kejahatan tertinggi yang terjadi di Indonesia pada tahun 2019-2022 adalah kasus pencurian dengan pemberatan. Dari

10 jenis kejahatan tertinggi di atas, 4 di antaranya terkategori sebagai kasus pencurian, yaitu kasus pencurian dengan pemberatan, pencurian biasa, curanmor R-2, dan pencurian dengan kekerasan.

Kota Palembang menjadi salah satu kota yang memiliki tingkat pencurian yang tergolong tinggi. Menurut Kapolrestabes Palembang pada tahun 2020, dari berbagai macam kejahatan, terdapat 4 kejahatan yang penyelesaian kasusnya menjadi prioritas dikarenakan kejahatan tersebut membuat resah masyarakat, yaitu pencurian dengan pemberatan, pencurian dengan kekerasan, pencurian kendaraan bermotor, dan penganiayaan berat. Pencurian adalah perbuatan mengambil benda atau harta milik orang lain tanpa seizin pemiliknya, tindakan pencurian merupakan salah satu tindak pidana yang sangat meresahkan masyarakat. Berdasarkan KUHP, tindak pidana pencurian terdiri dari lima jenis, yaitu pencurian biasa, pencurian ringan, pencurian dengan pemberatan, pencurian dengan kekerasan, dan pencurian dalam kalangan keluarga. Tindak pencurian dapat disebabkan oleh beberapa faktor, salah satunya adalah motivasi finansial, seseorang bisa saja melakukan pencurian dikarenakan membutuhkan uang untuk memenuhi kebutuhan hidup. Selain itu, pencurian bisa terjadi dikarenakan adanya kesempatan dan kurangnya pengawasan atau keamanan yang memadai. Agar masyarakat dapat memahami cara kerja para pelaku pencurian serta dapat mencegah tindak pidana pencurian di masa yang akan datang adalah dengan mengetahui karakteristik dari kasus pencurian yang telah terjadi. Dengan diketahuinya karakteristik pencurian, masyarakat dapat memahami bagaimana seorang pencuri beraksi, sehingga dapat lebih waspada dan mengambil tindak pencegahan agar menghindari dan meminimalisir risiko menjadi korban pencurian, dengan cara memasang kunci tambahan di rumah dan kendaraan serta memasang sistem keamanan yang lebih baik.

Berdasarkan uraian di atas, maka pada penelitian ini akan dilakukan pemodelan topik untuk mengidentifikasi karakteristik kasus tindak pidana pencurian pada kasus di Pengadilan Negeri Palembang, contohnya seperti daerah tempat terjadinya pencurian, jenis barang curian, target pencurian, alat pendukung pencurian, jenis pencurian, dan lain sebagainya. *Topic Modeling* bertujuan untuk menemukan topik dan kata yang tersembunyi pada kumpulan dokumen besar. Oleh karena itu, peneliti tertarik untuk mengidentifikasi karakteristik kasus tindak

pidana pencurian dari data surat dakwaan kasus tindak pidana pencurian di Pengadilan Negeri Palembang tahun 2019 -2022. Selain itu, dari hasil dari pemodelan topik antar tahun akan dilakukan perbandingan untuk mengetahui apakah terdapat perbedaan karakteristik kasus tindak pidana pencurian pada tahun 2019-2022, mengingat pada tahun tersebut merupakan masa-masa sebelum adanya Covid-19, saat merebaknya Covid-19, hingga masa pemulihan pasca terjadinya pandemi Covid-19. Terdapat beberapa metode untuk melakukan pemodelan topik, salah satunya adalah *Probabilistic Latent Semantic Analysis* (PLSA) dan *Latent Dirichlet Allocation* (LDA), namun, menurut Blei dalam jurnal (Setijohatmo et al., 2020) LDA dapat mengatasi permasalahan *overfitting* yang dialami metode PLSA. *Overfitting* merupakan suatu kondisi dimana model terdapat terlalu banyak parameter yang mengakibatkan tingkat kecocokan yang tinggi terhadap sampel, namun dengan sampel baru tingkat kecocokan tersebut akan menjadi rendah. Pada penelitian ini, metode yang peneliti gunakan untuk melakukan pemodelan topik adalah metode LDA. Menurut Xu et al dalam jurnal (Syarifuddin et al., 2020), LDA merupakan model probabilistik dalam *Topic Modeling* pada data *text* untuk mendapatkan informasi berupa model topik. Dengan dilakukannya analisis tersebut akan mempermudah dalam mengidentifikasi karakteristik kasus tindak pidana pencurian yang terjadi di Kota Palembang.

1.2. Rumusan Masalah

Berdasarkan latar belakang di atas, maka rumusan masalah yang dapat diidentifikasi adalah:

1. Bagaimana gambaran umum pada dokumen Berita Acara Kasus Tindak Pidana Pencurian di Pengadilan Negeri Palembang Tahun 2019-2022?
2. Bagaimana karakteristik kasus tindak pidana pencurian berdasarkan hasil implementasi *Topic Modeling* dengan metode *Latent Dirichlet Allocation* (LDA) pada dokumen Surat Dakwaan Kasus Tindak Pidana Pencurian di Pengadilan Negeri Palembang Tahun 2019-2022?
3. Bagaimana hasil perbandingan topik yang didapatkan pada tahun 2019 hingga 2022?

1.3. Batasan Masalah

Batasan masalah yang diberikan adalah sebagai berikut:

1. Data yang digunakan merupakan data sekunder yang diperoleh dari website Direktori Putusan Mahkamah Agung Republik Indonesia, yaitu <https://putusan3.mahkamahagung.go.id/> diakses dan diambil tanggal 20 Desember 2022 sampai dengan 9 Januari 2023.
2. Data yang diambil merupakan data yang dipublikasikan.
3. Data yang digunakan untuk analisis *Topic Modeling* adalah *text* yang memuat surat dakwaan penuntut umum.
4. Penelitian hanya difokuskan pada Surat Dakwaan Kasus Tindak Pidana Pencurian di Pengadilan Negeri Palembang Tahun 2019 hingga 2022.
5. Metode yang digunakan untuk analisis *Topic Modeling* adalah *Latent Dirichlet Allocation* (LDA).
6. Analisis data *text* diolah dengan menggunakan bahasa *Python* menggunakan *software Visual Studio Code*, dan menggunakan *software Microsoft Excel* menyimpan data serta pengolahan diagram.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Untuk mengetahui gambaran umum pada dokumen Berita Acara Kasus Tindak Pidana Pencurian di Pengadilan Negeri Palembang Tahun 2019-2022.
2. Untuk mengetahui karakteristik kasus tindak pidana pencurian berdasarkan hasil implementasi *Topic Modeling* dengan metode *Latent Dirichlet Allocation* (LDA) pada dokumen Surat Dakwaan Kasus Tindak Pidana Pencurian di Pengadilan Negeri Palembang Tahun 2019-2022.
3. Untuk mengetahui apakah terdapat perbedaan karakteristik kasus tindak pidana pencurian yang teridentifikasi berdasarkan hasil perbandingan topik tahun 2019-2022.

1.5. Manfaat Penelitian

1. Menambah pengetahuan dan wawasan mengenai analisis *Topic Modeling* dengan menggunakan metode *Latent Dirichlet Allocation* (LDA).

2. Dari penelitian ini, diharapkan dapat memberikan informasi mengenai Kasus Tindak Pidana Pencurian yang sering terjadi di Kota Palembang dan sekitarnya pada tahun 2019-2022 kepada penegak hukum maupun khalayak umum.
3. Dapat dijadikan referensi bagi para pembaca untuk penelitian selanjutnya.

BAB II

TINJAUAN PUSTAKA

Penelitian terdahulu dibutuhkan untuk menjadi bahan acuan dan dapat memberikan pemahaman terkait teori-teori yang berkaitan dengan topik serta metode yang digunakan dalam penelitian. Penelitian yang menjadi acuan merupakan penelitian terdahulu yang berkaitan analisis *Topic Modeling* pada *text* dokumen. Berikut ini penelitian-penelitian yang digunakan sebagai acuan dalam penelitian.

Penelitian yang dilakukan oleh Putri & Giri (2020) membahas mengenai topik perbincangan dalam jaringan media sosial perbankan. Penelitian ini bertujuan untuk mengetahui topik pembicaraan dalam jaringan sosial Bank BNI pada Twitter. Dari hasil penelitian disimpulkan bahwa topik pembicaraan dalam jaringan Bank BNI adalah mengenai promo terkait bank, terutama promo kuis berhadiah “#BNIQuotes” yang diadakan oleh akun resmi Bank BNI.

Penelitian yang dilakukan oleh Pranata et al (2020) bertujuan untuk mendapatkan model *Autoregressive* (AR), *Moving Average* (MA), dan model *Autoregressive Integrated Moving Average* (ARIMA) yang kemudian dilakukan pemilihan model terbaik untuk melakukan peramalan pada data pencurian sepeda motor. Diperoleh kesimpulan bahwa model AR, MA, dan ARIMA yang cocok pada kasus pencurian sepeda motor dengan pemberatan dan kekerasan adalah model MA(1) dengan nilai RMSE sebesar 6.5612926 dan nilai AIC sebesar 394.82.

Penelitian yang dilakukan oleh Naury et al (2021) menggunakan metode LDA dan LSTM untuk melakukan analisis *Topic Modeling* pada sentimen terhadap *headline* berita *online* berbahasa Indonesia. Analisis dilakukan dengan membuat model analisis sentimen terlebih dahulu, lalu melakukan analisis sentimen, kemudian dilanjutkan dengan proses pemodelan topik, serta visualisasi. Analisis sentimen dengan metode LSTM mendapatkan hasil akurasi tertinggi sebesar 71.13% dan terendah sebesar 63.48%, sedangkan hasil pemodelan topik berdasarkan hasil visualisasi yang disajikan didapatkan topik-topik yang memiliki *insight* dari suatu isu berita.

Penelitian yang dilakukan oleh Nasution et al (2021) membahas mengenai peristiwa politik yang membuat masyarakat Indonesia menyikapi dengan berbagai macam tanggapan sehingga akan dilakukan analisis dari algoritma LDA untuk menentukan topik politik pada twitter. Penelitian dilakukan dengan cara melakukan *preprocessing* data, kemudian pelatihan data, pengujian dan evaluasi, sehingga didapatkan hasil penelitian dengan pengujian 3 kali dan jumlah data masing-masing 100, 1000, dan 6000 dengan jumlah topik sebanyak 10 menghasilkan rata-rata nilai kebenaran 90%. Sehingga disimpulkan bahwa LDA dapat digunakan dan memiliki nilai kebenaran yang tinggi dalam mendeteksi topik politik pada Twitter.

Penelitian yang dilakukan oleh Salsabila & Trianasari (2021) membahas mengenai suatu produk kosmetik yaitu Laneige Water Sleeping Mask. Penelitian dilakukan dengan tujuan untuk mengetahui sentimen dan topik serta kata apa saja yang sering dibicarakan terkait dengan produk tersebut agar dapat digunakan perusahaan sebagai informasi untuk dapat terus meningkatkan kualitasnya di kemudian hari. Hasil penelitian menggambarkan ulasan yang didominasi oleh sentimen positif serta topik dan kata-kata yang mengungkapkan kepuasan konsumen terhadap produk Laneige Water Sleeping Mask.

Penelitian yang dilakukan Novianti et al (2021) bertujuan untuk menentukan tingkat kriminalitas kasus pencurian sepeda motor dengan penerapan metode Exponential Smoothing untuk menentukan tingkat kriminalitas pencurian sepedamotor di kota Tanjung Balai untuk periode berikutnya. Diperoleh hasil prediksi peramalan jumlah kasus pencurian sepeda motor yang terjadi di kota Tanjung Balai tahun 2020 adalah 12 unit dengan nilai error MAPE sebesar 0.153%.

Penelitian yang dilakukan oleh Ikegami & Darmawan (2022) merupakan pengamatan tren ulasan hotel dengan menggunakan metode *Latent Dirichlet Allocation*. Penelitian ini bertujuan mendapatkan gambaran dan ringkasan yang berupa kata-kata yang mengarah pada suatu topik dengan data yang digunakan adalah data ulasan hotel pada Tripadvisor dengan jumlah data sebanyak 20,491 baris yang terdiri dari 2 kolom yaitu *Review* dan *Rating*. Dari hasil pemodelan topik LDA yang telah dilakukan terhadap *dataset Review*, didapatkan kesimpulan bahwa tren ulasan lebih banyak membahas mengenai lokasi, pelayanan, hotel, sarapan, resort dan pantai.

Penelitian yang dilakukan oleh Utami & Suparyati (2022) menggunakan metode *Latent Dirichlet Allocation* untuk menentukan topik teks suatu berita. Penelitian ini bertujuan untuk memberikan solusi untuk masalah identifikasi topik pada artikel berita yang diterbitkan oleh situs berita *online* yaitu Detik.com. Penelitian ini menghasilkan 3 model topik yang hasil pemodelan divisualisasikan bentuk grafik untuk mengetahui trend topik yang sedang menjadi pemberitaan dengan hasil pengujian relevansi judul dengan topik menghasilkan akurasi 67%.

Penelitian yang dilakukan oleh Ardianti & Karmila (2022) mengombinasikan analisis sentimen dan pemodelan topik dalam memperoleh informasi terkait dengan pengalaman pengguna dalam menggunakan aplikasi Noice dengan menggunakan metode XGBoost dan LDA. Dari penelitian yang dilakukan diperoleh 6,286 ulasan merupakan sentimen positif dan 2,055 merupakan sentimen negatif. Dari hasil sentimen positif dan negatif dilakukan pemodelan topik, dari 3 topik yang didapatkan untuk ulasan dengan sentimen positif diketahui bahwa ulasan umumnya berisikan pujian terhadap aplikasi, konten yang disukai, serta alasan pengunduhan aplikasi oleh pengulas. Sedangkan, dari 3 topik yang didapatkan untuk ulasan sentimen negatif membahas mengenai permasalahan koneksi, pembukaan aplikasi, pemutaran atau pengunduhan konten, dan lain sebagainya.

Penelitian yang dilakukan oleh Fudholi & Ridhwanullah (2022) merupakan pemodelan topik dengan menggunakan metode *Latent Dirichlet Allocation* pada cuitan tentang penyakit tropis di Indonesia. Pada penelitian *coherence score* tertinggi yang didapatkan sebesar 0.576453 dengan jumlah topik yang didapatkan sebanyak 5. Dari hasil pemodelan topik dengan LDA, disimpulkan bahwa pembicaraan masyarakat Indonesia terkait dengan penyakit tropis seperti kusta, malaria, dan demam berdarah dengan media Twitter adalah mengenai kebutuhan dana untuk memerangi penyakit, Covid-19, kebutaan dan kusta, serta pengobatan dan penanggulangan.

Tabel 2.1 Tabel Penelitian Sebelumnya

No	Penulis	Judul Penelitian	Metode	Persamaan	Perbedaan
1.	Putri & Giri, 2020	Analisis <i>Topic Modeling</i> untuk Mengidentifikasi Topik	<i>Latent Dirichlet Allocation</i> (LDA)	Metode yang digunakan sama	Studi kasus yang digunakan berbeda

No	Penulis	Judul Penelitian	Metode	Persamaan	Perbedaan
		Pembicaraan pada Media Sosial Bank BNI			
2.	Pranata et al., 2020	Prediksi Pencurian Sepeda Motor Menggunakan Model <i>Time Seriuves</i> (Studi Kasus: Polres Kotabumi Lampung Utara)	<i>Autoregressive (AR), Moving Average (MA) & Autoregressive Integrated Moving Average</i>	Menggunakan data terkait kasus pencurian	Metode yang digunakan berbeda
3.	Naury et al., 2021	<i>Topic Modeling</i> pada Sentimen Terhadap <i>Headline Berita Online</i> Berbahasa Indonesia Menggunakan LDA dan LSTM	<i>Latent Dirichlet Allocation (LDA) & Long Short Term Memory (LSTM)</i>	Salah satu metode yang digunakan sama	Studi kasus yang digunakan berbeda
4.	Nasution et al., 2021	Sistem Deteksi Topik Politik pada Twitter Menggunakan Algoritma <i>Latent Dirichlet Allocation</i>	<i>Latent Dirichlet Allocation (LDA)</i>	Metode yang digunakan sama	Studi kasus yang digunakan berbeda
5.	Salsabila & Trianasari, 2021	Analisis Persepsi Produk Kosmetik Menggunakan Metode <i>Sentiment Analysis</i> dan <i>Topic Modeling</i> (Studi Kasus: Laneige Water Sleeping Mask)	<i>Latent Dirichlet Allocation (LDA)</i>	Metode yang digunakan sama	Studi kasus yang digunakan berbeda
6.	Novianti et al., 2021	Metode <i>Exponential Smoothing</i> pada Peramalan Tingkat Kriminalitas	<i>Exponential Smoothing</i>	Menggunakan data terkait kasus pencurian	Metode yang digunakan berbeda

No	Penulis	Judul Penelitian	Metode	Persamaan	Perbedaan
		Pencurian Sepeda Motor			
7.	Ikegami & Darmawan, 2022	Pengamatan Tren Ulasan Hotel Menggunakan Pemodelan Topik Berbasis <i>Latent Dirichlet Allocation</i>	<i>Latent Dirichlet Allocation</i> (LDA)	Metode yang digunakan sama	Studi kasus yang digunakan berbeda
8.	Ardianti & Karmila, 2022	Metode <i>Latent Dirichlet Allocation</i> untuk Menentukan Topik Teks Suatu Berita	<i>Latent Dirichlet Allocation</i> (LDA)	Metode yang digunakan sama	Studi kasus yang digunakan berbeda
9.	Utami & Suparyati, 2022	Analisis Sentimen dan Pemodelan Topik Ulasan Aplikasi Noice Menggunakan XGBoost dan LDA	<i>XGBoost</i> dan <i>Latent Dirichlet Allocation</i> (LDA)	Salah satu metode yang digunakan sama	Studi kasus yang digunakan berbeda
10.	Fudholi & Ridhwanullah, 2022	Pemodelan Topik pada Cuitan tentang Penyakit Tropis di Indonesia dengan Metode <i>Latent Dirichlet Allocation</i>	<i>Latent Dirichlet Allocation</i> (LDA)	Metode yang digunakan sama	Studi kasus yang digunakan berbeda

Pada penelitian terdahulu yang menggunakan studi kasus terkait kasus pencurian hanya melakukan peramalan saja, dengan menggunakan model *time series* maupun *exponential smoothing*. Sedangkan pada penelitian ini, peneliti melakukan pemodelan topik sehingga akan memberikan informasi mengenai karakteristik pencuriannya, seperti dimana daerah yang banyak terjadi kasus pencurian, jenis barang yang dicuri, target pencurian, dan lain sebagainya.

BAB III

LANDASAN TEORI

3.1. Tindak Pidana

Tindak pidana merupakan pengertian dasar dalam hukum pidana, tindak pidana merupakan suatu pengertian yuridis, lain halnya dengan istilah perbuatan jahat atau kejahatan. Istilah tindak pidana berasal dari istilah yang dikenal dalam hukum Belanda yaitu “*strafbaar feit*”. *Strafbaar feit* terdiri dari tiga kata yakni *straf*, *baar*, dan *feit*, yang mana ketika *straf* diterjemahkan maka akan sama dengan pidana dan hukum, sedangkan kata *baar* diterjemahkan dengan istilah dapat atau boleh. Kata *feit* diterjemahkan dengan tindak, peristiwa, pelanggaran dan perbuatan (Rahim, 2019).

Tindak pidana adalah suatu tindakan yang dilakukan oleh seseorang yang melanggar peraturan perundang-undangan yang berlaku sehingga dirinya dipersalahkan atas perbuatan tersebut dan menimbulkan kerugian harta benda ataupun yang bukan harta benda dan disertai dengan ancaman sanksi pidana dan hukum yang relevan (Suari et al., 2020). Perbuatan pidana adalah perbuatan yang dilarang oleh Undang-Undang yang disertai dengan sanksi yang berupa hukuman tertentu bagi siapapun yang melanggar larangan tersebut. Dapat juga dikatakan bahwa perbuatan pidana adalah perbuatan yang oleh suatu aturan hukum dilarang dan diancam pidana, namun perlu diingat bahwa larangan ditujukan kepada perbuatan (yaitu suatu keadaan atau kejadian yang ditimbulkan oleh tingkah laku seseorang), sedangkan ancaman pidananya ditunjukkan kepada orang yang menimbulkan kejadian itu (Sari, 2020).

3.2. Pencurian

Menurut Fauzi & Jainah (2022) pencurian adalah tindak pidana yang ditunjukkan terhadap harta benda seseorang. Kejahatan jenis ini sering terjadi di tengah-tengah masyarakat, meskipun perbuatan ini bukan termasuk kedalam golongan yang sangat berat seperti pembunuhan, pencabulan, dan korupsi, namun tetap dapat menimbulkan keresahan di masyarakat, terutama di lingkungan tempat tinggal itu sendiri.

Sedangkan, menurut Zazin et al (2022) pencurian adalah salah satu kejahatan yang merupakan suatu tindakan mengambil barang milik orang lain secara melawan hukum. Tindak pidana pencurian ini tergolong menjadi pencurian biasa, ringan, pemberatan, bahkan dengan kekerasan yang mana bisa hampir setiap hari terlihat di media elektronik dan media massa.

Menurut Saputra (2019) tindak pidana pencurian merupakan fenomena sosial yang selalu tampak dalam kehidupan bermasyarakat, telah dilakukan berbagai upaya untuk menghilangkannya, baik oleh pihak berwajib maupun warga masyarakat sendiri, akan tetapi upaya tersebut tidak dapat terwujud sepenuhnya, karena setiap kejahatan tidak akan dihapuskan dengan mudah melainkan hanya dapat dikurangi tingkat intensitasnya dan kualitasnya.

Berdasarkan Kitab Undang-Undang Hukum Pidana, tindak pidana pencurian terdiri dari lima jenis, yaitu pencurian biasa, pencurian ringan, pencurian dengan pemberatan, pencurian dengan kekerasan, dan pencurian dalam kalangan keluarga (Suryadi & Supriatna, 2019).

3.2.1 Pencurian Biasa

Istilah “pencurian biasa” digunakan oleh pakar hukum pidana untuk menunjuk pengertian “pencurian dalam arti pokok”. Perumusan pencurian biasa diatur dalam Pasal 362 KUHP yang menyatakan: Barangsiapa mengambil sesuatu barang, yang sama sekali atau sebagian termasuk kepunyaan orang lain, dengan maksud akan memiliki barang itu dengan melawan hak, dihukum, karena pencurian, dengan hukuman penjara selama-lamanya lima tahun atau denda sebanyak-banyaknya Rp. 900,-

3.2.2 Pencurian Ringan

Pencurian ringan diatur dalam Pasal 364 KUHP yang menyatakan: Perbuatan yang diterangkan dalam Pasal 362 dan Pasal 363 Ayat 4 begitu juga perbuatan yang diterangkan dalam Pasal 365 Ayat 5, apabila tidak dilakukan dalam sebuah rumah atau pekarangan tertutup yang ada rumahnya, jika harga barang yang dicuri tidak lebih dari dua puluh lima ribu rupiah, dihukum sebagai pencurian ringan, pidana penjara paling lama tiga bulan atau denda paling banyak Sembilan Ratus Rupiah.

3.2.3 Pencurian dengan Pemberatan

Suatu perbuatan dapat digolongkan sebagai pencurian dengan pemberatan apabila memenuhi unsur-unsur dalam Pasal 362 KUHP, juga harus memenuhi unsur lain yang terdapat dalam Pasal 363 KUHP.

3.2.4 Pencurian dengan Kekerasan

Pencurian dengan pemberatan kedua adalah pencurian yang diatur dalam Pasal 365 KUHP. Jenis pencurian ini lazim disebut dengan istilah “pencurian dengan kekerasan” atau populer dengan istilah “curas”

3.2.5 Pencurian Dalam Kalangan Keluarga

Pencurian dikalangan keluarga diatur dalam Pasal 367 KUHP, artinya, baik pelaku maupun korbannya masih dalam satu keluarga, misalnya yang terjadi, apabila seorang suami atau istri melakukan (sendiri) atau membantu (orang lain) pencurian terhadap harta benda istri atau suaminya.

3.3. Surat Dakwaan

Surat dakwaan adalah surat yang berisi dakwaan bahwa seseorang telah melakukan tindak pidana. Pada dasarnya, surat dakwaan adalah kesimpulan Jaksa Penuntut Umum tentang apa yang dilakukan oleh tersangka, berdasarkan hasil penyidikan dan dasar bagi penuntut umum dalam mengajukan terdakwa ke pengadilan. Surat dakwaan adalah dasar bagi pemeriksaan perkara selanjutnya, baik pemeriksaan di persidangan Pengadilan Negeri, pemeriksaan di tingkat banding dan kasasi. Dengan kata lain, Jaksa Penuntut Umum hanya dapat menuntut terdakwa tidak boleh lebih dari ancaman yang dikenakan dalam Pasal surat dakwaan (Hutapea et al., 2020).

3.4. Statistika Deskriptif

Statistika deskriptif adalah cabang statistika yang membahas cara pengumpulan dan penyajian data sehingga mudah dipahami dan menghasilkan informasi yang bermanfaat. Fungsi dari statistika deskriptif yaitu menggambarkan dan menjelaskan situasi dan persoalan tanpa menarik suatu kesimpulan (Hutasuhut, 2022).

Statistika deskriptif adalah statistik yang digunakan untuk menganalisis data dengan cara mendeskripsikan atau menggambarkan data yang telah terkumpul

sebagaimana adanya tanpa bermaksud menarik kesimpulan umum atau generalisasi (Rumahorbo et al., 2022).

3.5. Data Mining

Data mining adalah kegiatan mengekstraksi atau penggalian informasi dari data yang berukuran atau berjumlah besar, informasi inilah yang nantinya sangat berguna untuk pengembangan. Definisi sederhana dari *data mining* adalah ekstraksi informasi atau pola yang penting atau menarik dari data yang ada di *database* yang besar. Dalam jurnal ilmiah, *data mining* juga dikenal dengan nama *Knowledge Discovery in Databases* (KDD). Alasan utama mengapa *data mining* sangat menarik perhatian industri informasi dalam beberapa tahun terakhir adalah karena ketersediaan data dalam jumlah yang besar dan kebutuhan yang terus meningkat untuk mengubah data tersebut menjadi informasi dan pengetahuan yang berguna (Susanto & Sudiyatno, 2014).

Data mining adalah proses pencarian pola atau informasi menarik pada data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode, atau algoritma dalam *data mining* sangat bervariasi. Memilih metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database* (KDD) secara keseluruhan (Mardi, 2017).

3.6. Text Mining

Text mining adalah sebuah teknik untuk menggali informasi dari sebuah teks. *Text mining*, mengacu pada proses mengambil informasi berkualitas tinggi dari teks. Informasi berkualitas tinggi biasanya diperoleh melalui peramalan pola dan kecenderungan melalui sarana seperti pembelajaran pola statistik. Proses *text mining* yang khas meliputi :

1. Kategorisasi teks,
2. *Text clustering*, ekstraksi konsep/ entitas,
3. Produksi taksonomi granular,
4. *Sentiment analysis*, penyimpulan dokumen, dan
5. Pemodelan relasi entitas (yaitu, pembelajaran hubungan antara entitas bernama) (Susanto & Sudiyatno, 2014).

3.7. *Preprocessing*

Preprocessing adalah tahap proses awal *text mining* terhadap teks untuk mempersiapkan teks menjadi data yang dapat diolah lebih lanjut (Mansyur et al., 2019). Tahap *preprocessing* yang peneliti lakukan adalah *case folding*, *remove punctuation*, *remove number*, *stopword removal*, *stemming* dan *tokenizing*.

1. *Case folding* merupakan proses mengubah setiap karakter huruf dalam isi dokumen menjadi huruf kecil. Tidak semua dokumen teks konsisten dengan penggunaan huruf besar-kecil. Oleh karena itu, peran *case folding* dibutuhkan dalam mengonversi semua teks dalam dokumen menjadi format standar (biasanya huruf kecil) (Adi, 2018).
2. *Remove punctuation* adalah penghapusan semua tanda baca selain alphabet pada teks.
3. *Remove number* adalah penghapusan semua angka yang terdapat pada teks (Fani, Santoso, & Suparti, 2021).
4. *Stopword removal* merupakan proses penghilangan kata tertentu dari sebuah data teks yang dianggap tidak relevan (*stopword*). Pada dasarnya, *stopword* adalah daftar kata dalam suatu bahasa. *Stopword* cenderung dihilangkan pada penelitian terkait *text mining*, karena penggunaan *stopword* yang berulang-ulang dalam sebuah kalimat, sehingga *stopword* dihilangkan agar penelitian dapat lebih fokus pada kata-kata yang lebih penting (Risnantoyo et al., 2020). Contoh *stopword* pada teks Bahasa Indonesia yaitu, di, ke, dan, pada, dan lain sebagainya.
5. *Stemming* merupakan proses mengubah kata menjadi kata dasar dengan menghilangkan imbuhan-imbuhan pada kata dalam dokumen atau mengubah kata kerja menjadi kata benda (Gozali et al., 2020).
6. *Tokenizing* adalah tahap pemotongan *string input*-an berdasarkan bagian-bagian penyusunnya, atau dengan kata lain, kalimat dipecah menjadi kata-kata. Strategi umum yang digunakan dalam tahapan ini adalah memotong kata dengan spasi dan menghapus tanda baca. Tahap *tokenizing* membagi urutan karakter menjadi kalimat dan kalimat menjadi token (Jumeilah, 2017).

3.8. *Term Frequency-Inverse Document Frequency (TF-IDF)*

Pembobotan kata adalah proses pemberian bobot pada setiap kata yang terdapat dalam sebuah dokumen. Salah satu metode paling populer untuk mencari informasi peringkat berdasarkan frekuensi kata adalah metode TF-IDF (*Term Frequency - Inverse Document Frequency*) (Gunawan et al., 2018). Pendekatan dalam pembobotan kata yang paling banyak diterapkan adalah *term frequency* (TF). Faktor ini menunjukkan jumlah kemunculan kata dalam suatu dokumen. Semakin sering suatu kata muncul dalam dokumen, maka semakin penting kata tersebut. Terdapat empat cara yang bisa digunakan untuk mendapatkan nilai TF:

1. *Raw TF*

Nilai TF suatu *term* dihitung berdasarkan kemunculan *term* tersebut dalam dokumen.

2. *Logarithmic TF*

Cara ini menggunakan fungsi logaritmik dalam matematika untuk mendapatkan nilai TF.

$$TF = 1 + \log(TF) \quad (3.1)$$

3. *Binnary TF*

Cara ini menghasilkan nilai *Boolean* berdasarkan kemunculan *term* dalam dokumen. Akan bernilai 0 apabila *term* tidak ada dalam dokumen, dan bernilai 1 apabila *term* tersebut ada dalam dokumen. Dengan demikian, jumlah kemunculan suatu *term* dalam suatu dokumen tidak berpengaruh.

4. *Augmented TF*

$$TF = 0.5 + 0.5 \times TF_{max}(TF) \quad (3.2)$$

Nilai TF adalah jumlah kemunculan *term* dalam dokumen. Nilai *max* (TF) adalah jumlah kemunculan terbanyak *term* pada dokumen yang sama. (Siregar et al., 2017).

Inverse Document Frequency (IDF) menunjukkan hubungan ketersediaan sebuah *term* dalam seluruh dokumen. Semakin rendah nilai TF, maka akan semakin tinggi nilai IDF (Utomo et al., 2019). IDF digunakan untuk menilai keunikan kata dalam dokumen, kata yang muncul dalam jumlah dokumen yang lebih sedikit akan dianggap lebih penting atau memiliki bobot yang lebih besar. Berikut merupakan persamaan untuk menghitung IDF.

$$IDF_{(t)} = \ln\left(\frac{D}{DF_{(t)}}\right) \quad (3.3)$$

dimana:

$IDF_{(t)}$: nilai *Inverse Document Frequency* suatu *term* (t)

D : jumlah keseluruhan dokumen

$DF_{(t)}$: jumlah dokumen yang mengandung *term* (t)

Term Frequency – Inverse Document Frequency (TF-IDF) merupakan algoritma yang digunakan untuk menghitung bobot setiap kata pada masing-masing dokumen terhadap kata kunci. *Term Frequency* merupakan frekuensi kemunculan suatu kata dalam dokumen yang bersangkutan, sedangkan *Inverse Document Frequency* (IDF) adalah suatu statistik yang mengkarakteristikan sebuah kata dalam keseluruhan dokumen. *Document Frequency* (DF) adalah jumlah dokumen dimana kata tertentu muncul. TF merupakan pembobotan yang sederhana dimana tingkat kepentingan suatu kata diasumsikan sebanding dengan jumlah kemunculan kata tersebut dalam dokumen, sementara IDF adalah pembobotan yang mengukur tingkat kepentingan suatu kata dalam dokumen. Nilai pembobotan TF-IDF akan tinggi jika nilai TF besar dan kata yang diamati tidak ditemukan dalam banyak dokumen (Laxmi et al., 2019). Oleh karena itu, untuk mencari nilai TF-IDF menggunakan **Persamaan 3.4**.

$$w_{(t,d)} = TF_{t,d} \times IDF_t \quad (3.4)$$

dimana:

$w_{(t,d)}$: bobot dari suatu kata (t) dalam satu dokumen

$TF_{t,d}$: frekuensi kemunculan suatu kata (t) dalam dokumen d

IDF_t : nilai *Inverse Document Frequency* suatu *term* (t)

Jika frekuensi *term* sama dengan jumlah dokumen, maka hasil perhitungan $IDF_t = 0$. Untuk menghindari hasil $w_{(t,d)} = 0$, dapat ditambahkan dengan nilai 1 pada hasil perhitungan IDF (Apriani et al., 2021), seperti pada **Persamaan 3.5**.

$$w_{(t,d)} = TF_{t,d} \times (IDF_t + 1) \quad (3.5)$$

3.9. Topic Modeling

Topic Modeling terdiri dari entitas-entitas yaitu “kata”, “dokumen”, dan “*corpus*”. “Kata” dianggap sebagai unit dasar dari data diskrit dalam dokumen, yang didefinisikan sebagai item dari kosakata yang diberi indeks untuk setiap kata unik yang ada dalam dokumen. “Dokumen” merupakan susunan N kata-kata. Sebuah *corpus* adalah kumpulan M dokumen. “*Topic*” adalah distribusi dari beberapa kosakata yang bersifat tetap. Secara sederhana, setiap dokumen dalam *corpus* mengandung proporsi yang berbeda dari topik-topik yang dibahas sesuai dengan kata-kata yang ada di dalamnya.

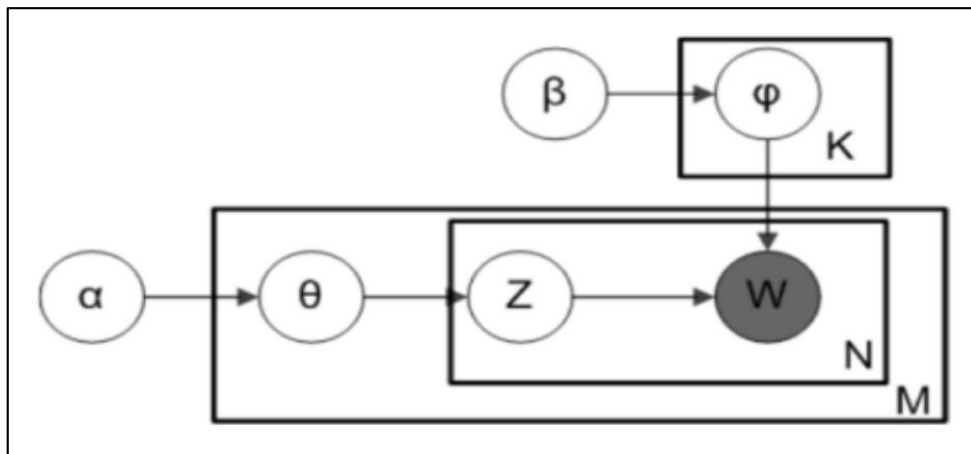
Ide dasar dari *Topic Modeling* ialah sebuah topik terdiri dari kata-kata tertentu yang membentuk topik tersebut, dan dalam satu dokumen memiliki kemungkinan terdiri dari beberapa topik dengan probabilitas masing-masing. Namun, manusia memahami dokumen-dokumen merupakan sebuah objek yang dapat diamati, sedangkan topik, distribusi topik per-dokumen, dan penggolongan setiap kata pada topik per dokumen merupakan bagian yang tersembunyi. Oleh karena itu, pemodelan topik bertujuan untuk menemukan topik dan kata yang tersembunyi dalam suatu dokumen.

Dengan distribusi probabilitas topik di setiap dokumen, dapat diketahui seberapa banyak masing-masing topik terlibat dalam sebuah dokumen. Hal ini dapat mengetahui topik mana yang dibicarakan suatu dokumen (Rashif et al., 2021).

3.10. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) adalah algoritma pemodelan topik probabilistik yang bersifat generatif. Algoritma ini digunakan sebagai pembelajaran tanpa pengawasan untuk data yang tidak terstruktur. LDA dapat menemukan dan mengkategorikan kata kunci dalam dokumen ke dalam masing-masing topik. Metode ini didasarkan pada gagasan bahwa dokumen terdiri dari campuran topik yang tersembunyi dan topik-topik tersebut dicirikan oleh distribusi kata. *Latent* didefinisikan sebagai sesuatu yang ada tetapi tidak terlihat. LDA sendiri bekerja dengan *Gibbs Sampling*, yaitu menghitung distribusi probabilitas secara bersama-sama dengan cara mengambil sampel setiap variabel satu per satu berdasarkan nilai dari variabel lainnya (Dikiyanti et al., 2021).

Menurut Natalia et al (2021), dalam penerapannya LDA membentuk sebuah sistem untuk menentukan beberapa kelompok topik berdasarkan data atau informasi yang terdiri dari kata-kata. Metode LDA ini merupakan metode improvisasi dari dua metode yang sebelumnya telah diperkenalkan terlebih dahulu yakni metode *Probabilistic Latent Semantic Analysis* (PLSA) dan *Latent Semantic Analysis* (LSA). **Gambar 3.1** merupakan gambaran dari *graphical model* LDA menggunakan *plate notation*.



Gambar 3.1 Plate Notation LDA (Sumber: Setijohatmo et al., 2020)

Berikut ini merupakan formulasi dari *Latent Dirichlet Allocation*:

$$p(w, z, \theta, \varphi | \alpha, \beta) = p(\varphi | \beta) p(\theta | \alpha) p(z | \theta) p(w | \varphi_k) \quad (3.6)$$

Dari **Gambar 3.1** yang tersaji di atas dapat diketahui bahwa metode LDA memiliki beberapa variabel yaitu:

α : distribusi topik dalam dokumen pada suatu *corpus*

θ : probabilitas i dokumen yang mengandung j topik

β : probabilitas i topik yang mengandung j kata

M : jumlah dokumen

N : jumlah kata dalam setiap dokumen

w : kata dalam dokumen

z : topik yang terbentuk

φ : distribusi kata terhadap topik dalam *corpus*

Menurut Prihatini et al dalam jurnal (Syaifuddin et al., 2020) menyatakan bahwa algoritma LDA terdiri dari proses inisialisasi, proses berulang dan *sampling*, dan proses membaca parameter akhir.

1. Tahap inisialisasi merupakan proses penentuan frekuensi kemunculan kata pada setiap file teks. Proses ini dilakukan pada teks hasil *preprocessing* data. Proses inisialisasi dilakukan dengan langkah:
 - a. Menentukan indeks dari setiap kata dalam dokumen
 - b. Menghitung frekuensi kemunculan setiap kata pada setiap dokumen menggunakan *Bag-of-Words* (BoW)
 - c. Menentukan topik setiap kata dengan random berdasarkan nilai frekuensi kemunculan kata (z_0), LDA membutuhkan nilai topik yang ditentukan terlebih dahulu. Selain itu, dalam algoritma LDA, tidak ada nilai awal yang diberikan untuk setiap kata dalam dokumen, sehingga setiap kata memiliki nilai ketidakpastian.
 - d. Menentukan matriks kata-topik dan dokumen-topik
 - e. Menghitung jumlah total dari distribusi kata-topik dan dokumen-topik, dan menyimpan hasil matriks. Distribusi kata, W_i , pada tiap topik, Z_i , NW dilihat pada persamaan berikut

$$NW = \begin{Bmatrix} NW_{w1,z1} & NW_{w1,z2} & \dots & NW_{w1,zk} \\ NW_{w2,z1} & NW_{w2,z2} & \dots & NW_{w2,zk} \\ \vdots & \vdots & \ddots & \vdots \\ NW_{wn,z1} & NW_{wn,z2} & \dots & NW_{wn,zk} \end{Bmatrix} \quad (3.7)$$

dimana:

wn : *term* ke- n pada vocab

zk : topik ke- k

$NW_{wn,zk}$: banyak *term* ke- n yang berlabel topik ke- k

Kemudian membuat matriks distribusi topik, Z_k , pada tiap dokumen:

$$ND = \begin{Bmatrix} ND_{d1,z1} & ND_{d1,z2} & \dots & ND_{d1,zk} \\ ND_{d2,z1} & ND_{d2,z2} & \dots & ND_{d2,zk} \\ \vdots & \vdots & \ddots & \vdots \\ ND_{dm,z1} & ND_{dm,z2} & \dots & ND_{dm,zk} \end{Bmatrix} \quad (3.8)$$

dimana,

d_m : dokumen ke- m

z_k : topik ke- k

$ND_{dn,zk}$: banyak label topik ke- k pada dokumen ke- n

Pada akhir proses inisialisasi, dilakukan proses perhitungan jumlah setiap topik dalam dokumen (ND) dan jumlah setiap kata dalam topik

(NW) yang akan digunakan dalam proses iterasi dan *sampling* topik. Nilai total semua ND juga dihitung sebagai $SUMND$ dan nilai total semua NW sebagai $SUMNW$. Nilai dari $SUMND$ dan $SUMNW$ digunakan untuk mengurangi dan menambahkan nilai ND dan NW di sembarang perubahan topik yang terjadi pada setiap kata. Jumlah distribusi NW dan ND sebagai berikut:

$$NWSum_{zi} = \sum_{j=1}^m NW_{wj,zi} \dots \dots \dots \quad (3.9)$$

$$NWSum_{di} = \sum_{j=1}^k NW_{dj,zi} \dots \dots \dots \quad (3.10)$$

dimana,

$NWSum_{zi}$: jumlah seluruh kata dalam setiap topik

$NWSum_{di}$: jumlah seluruh topik dalam setiap dokumen

$NW_{wj,zi}$: jumlah setiap kata dalam topik

$NW_{dj,zi}$: setiap topik dalam dokumen

2. Tahap *sampling* topik merupakan proses penentuan topik baru dari setiap kata pada setiap dokumen. Proses ini dilakukan pada teks hasil *preprocessing*. Proses *sampling* topik dilakukan dengan langkah:

a. Menghitung probabilitas kata pada topik

$$\phi_{ij} = \frac{C_{ij}^{WT} + \beta}{\sum_{k=1}^W C_{kj}^{WT} + W\beta} \quad (3.11)$$

dimana,

ϕ_{ij} : probabilitas dari kata i untuk topik j

C_{ij}^{WT} : jumlah kata i pada topik j

WT : kata-topik

β : parameter sebaran *Dirichlet* distribusi kata terhadap topik

$\sum_{k=1}^W C_{kj}^{WT}$: jumlah seluruh kata pada topik j

k : indeks topik

W : jumlah seluruh kata pada dokumen

Menghitung probabilitas dokumen pada topik:

$$\theta_{dj} = \frac{C_{dj}^{DT} + \alpha}{\sum_{k=1}^T C_{dk}^{DT} + T\alpha} \quad (3.12)$$

dimana,

- θ_{dj} : proporsi topik j dalam dokumen d
- C_{dj}^{DT} : jumlah topik j pada dokumen d
- DT : dokumen-topik
- α : parameter sebaran *Dirichlet* distribusi topik terhadap dokumen
- $\sum_{k=1}^T C_{dk}^{DT}$: jumlah seluruh topik pada dokumen d
- T : jumlah seluruh topik yang sudah ditentukan

- b. Menentukan topik baru dari setiap kata dengan distribusi multinomial (posterior) berdasarkan nilai probabilitas kata tertinggi.

Parameter distribusi posterior dimana bobot tersebut diperoleh dari nilai distribusi probabilitas kata-topik dikali distribusi probabilitas topik-dokumen.

$$P(z_i = j | z_i, w_i, d_i) = \frac{C_{ij}^{WT} + \beta}{\sum_{k=1}^W C_{kj}^{WT} + W\beta} \times \frac{C_{dj}^{DT} + \alpha}{\sum_{k=1}^T C_{dk}^{DT} + T\alpha} \quad (3.13)$$

- c. Menyimpan hasil distribusi posterior
 - d. Langkah-langkah ini dilakukan sebanyak n perulangan sampai mencapai kondisi konvergen.
3. Tahap perhitungan parameter final merupakan proses untuk menghitung jumlah dokumen per topik dan jumlah kata per topik berdasarkan matriks kata-topik dan dokumen-topik yang telah konvergen.

3.11. Topic Coherence

Topic coherence digunakan untuk mengevaluasi model topik. *Topic coherence* didefinisikan sebagai rata-rata dari skor kemiripan kata berpasangan dari kata-kata dalam topik. Setiap topik dalam model terdiri dari kata-kata dan *topic coherence* diterapkan pada N kata teratas (Gangadharan & Gupta, 2020). *Coherence score* adalah metrik kinerja yang banyak digunakan untuk mengevaluasi metode pemodelan topik. Nilai *coherence* memberikan ukuran yang realistis untuk mengidentifikasi jumlah topik yang ada dalam dokumen. Setiap topik yang dihasilkan memiliki daftar kata seperti kelompok kata. Ukuran ini menemukan rata-

rata skor kemiripan dari kata-kata yang terkait dengan topik. Model yang baik adalah model yang menghasilkan topik dengan nilai *coherence* yang tinggi (George & Srividhya, 2020). Formulasi untuk menghitung nilai *coherene* ditunjukkan pada **Persamaan 3.14**.

$$coherence(V) = \sum_{(v_i, v_j) \in V} score(v_i, v_j, \epsilon) \quad (3.14)$$

dimana,

V : sekumpulan kata yang menggambarkan topik

ϵ : faktor penghalus yang menjamin bahwa skor menghasilkan bilangan riil

Fungsi berikut digunakan untuk menghitung seberapa sering dua kata v_i dan v_j , muncul bersamaan dalam teks.

$$score(v_i, v_j, \epsilon) = \log \frac{D(v_i, v_j)}{D(v_j)} \quad (3.15)$$

$D(v_i, v_j)$: menghitung jumlah dokumen yang mengandung kata-kata v_i dan v_j

$D(v_j)$: menghitung jumlah dokumen yang mengandung kata-kata v_j

(Tresnasari et al., 2020)

3.12. *Wordcloud*

Wordcloud adalah salah satu metode analisis dalam *text mining* yang dapat memberi visualisasi data teks yang menggambarkan frekuensi kata-kata yang ditampilkan dalam bentuk yang menarik namun informatif. Semakin besar ukuran kata pada tampilan *wordcloud* maka semakin sering kata tersebut digunakan (Cikania, 2021).

Wordcloud mengungkapkan seberapa sering kata-kata yang berbeda muncul dalam sebuah teks. Sampai batas tertentu, pemahaman tentang komposisi umum dari kata-kata yang sering digunakan memungkinkan pemirsa untuk memiliki gambaran umum tentang topik utama dan tema utama dalam sebuah teks, dan dapat menggambarkan sudut pandang utama yang dipegang oleh penulis teks. Perbandingan *wordcloud* yang dihasilkan dari teks yang berbeda akan dengan cepat mengungkapkan perbedaan antara ide-ide yang terkandung dalam teks-teks ini. (McNaught & Lam, 2010)

3.13. *Principal Component Analysis*

Principal Component Analysis merupakan suatu teknik yang digunakan untuk menyederhanakan data dengan cara mengubah data secara linier untuk

membentuk sistem koordinat baru dengan varians maksimum (M. T. Utami et al., 2017).

Menurut Smith dalam jurnal (Rahmawati, 2014) tujuan dari prosedur PCA adalah menyederhanakan variabel yang diamati dengan cara mereduksi dimensinya. Berikut merupakan tahapan mereduksi dimensi dalam PCA:

1. Menghilangkan korelasi diantara variabel *independent* dengan melakukan transformasi variabel *independent* asal menjadi variabel baru yang tidak berkorelasi sama sekali atau sering disebut dengan *principal component*.
2. Menghapus *eigenvector* serta *eigenvalue* yang sangat kecil sehingga dimensi akan berkurang dan tidak akan membuat kehilangan data yang penting. Hal tersebut dikarenakan, *eigenvector* dengan *eigenvalue* yang besar berperan paling penting dalam proses transformasi.

BAB IV

METODOLOGI PENELITIAN

Secara umum pada metodologi penelitian meliputi unsur-unsur:

4.1. Populasi Penelitian

Populasi yang digunakan dalam penelitian ini adalah dokumen Berita Acara Kasus Tindak Pidana Pencurian di Pengadilan Negeri Palembang. Sedangkan, sampel yang digunakan dalam penelitian adalah dokumen Berita Acara Kasus Tindak Pidana Pencurian di Pengadilan Negeri Palembang tahun 2019 – tahun 2022.

4.2. Jenis dan Sumber Data

Jenis data yang digunakan oleh peneliti merupakan data sekunder, yaitu data yang diperoleh dari *website* Direktori Putusan Mahkamah Agung Republik Indonesia <https://putusan3.mahkamahagung.go.id/>, diakses dan diambil tanggal 20 Desember 2022 sampai dengan 9 Januari 2023. Data yang digunakan merupakan dokumen yang terpublikasi pada *website* tersebut.

4.3. Variabel penelitian

Variabel yang digunakan dalam penelitian ini adalah Surat Dakwaan, Jenis Kelamin Terdakwa, Waktu Pencurian, dan Hukuman pada Kasus Tindak Pidana di Pengadilan Negeri Palembang tahun 2019 – 2022.

Tabel 4.1 Variabel Penelitian

No.	Variabel	Definisi Operasional Variabel
1	Surat Dakwaan	Surat dakwaan merupakan surat resmi yang dibuat oleh Jaksa Penuntut Umum, berisi dakwaan, rincian fakta-fakta yang mendukung tuduhan terhadap terdakwa dan bukti-bukti yang diajukan dalam sidang pengadilan.
2	Jenis Kelamin	Jenis kelamin para terdakwa
3	Waktu Pencurian	Waktu terjadinya pencurian
4	Hukuman	Lama hukuman yang ditetapkan kepada terdakwa

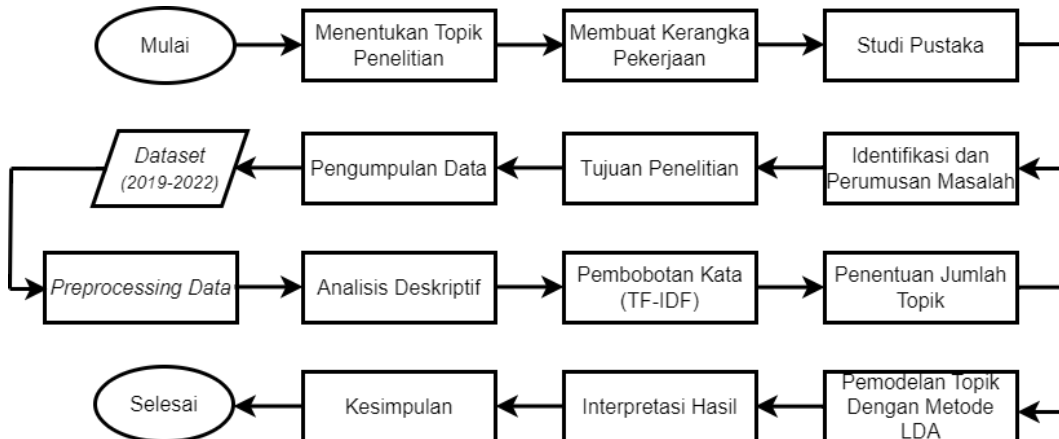
4.4. Metode dan Analisis Data

Software yang digunakan pada penelitian ini adalah *software Visual Studio Code*, dan menggunakan *software Microsoft Excel*. Berikut merupakan metode analisis data yang digunakan.

1. Analisis deskriptif yang digunakan untuk mengetahui gambaran umum pada data Berita Acara Kasus Tindak Pidana Pencurian di Pengadilan Negeri Palembang Tahun 2019-2022.
2. *Preprocessing* data yang dilakukan untuk menghapus atau memperbaiki kata agar data lebih terstruktur dan dapat mewakili dokumen, sehingga dapat dipahami oleh komputer.
3. Pembobotan kata yang digunakan untuk mengubah data *text* menjadi numerik dengan menggunakan metode TF-IDF.
4. Penentuan jumlah topik dengan menggunakan *topic coherence*.
5. Pemodelan topik dengan metode LDA untuk mengetahui karakteristik kasus tindak pidana pencurian di Pengadilan Negeri Palembang Tahun 2019 – 2022.

4.5. Tahapan Penelitian

Berikut ini merupakan tahapan-tahapan yang dilakukan dalam penelitian ini:



Gambar 4.1 Diagram Alir Penelitian

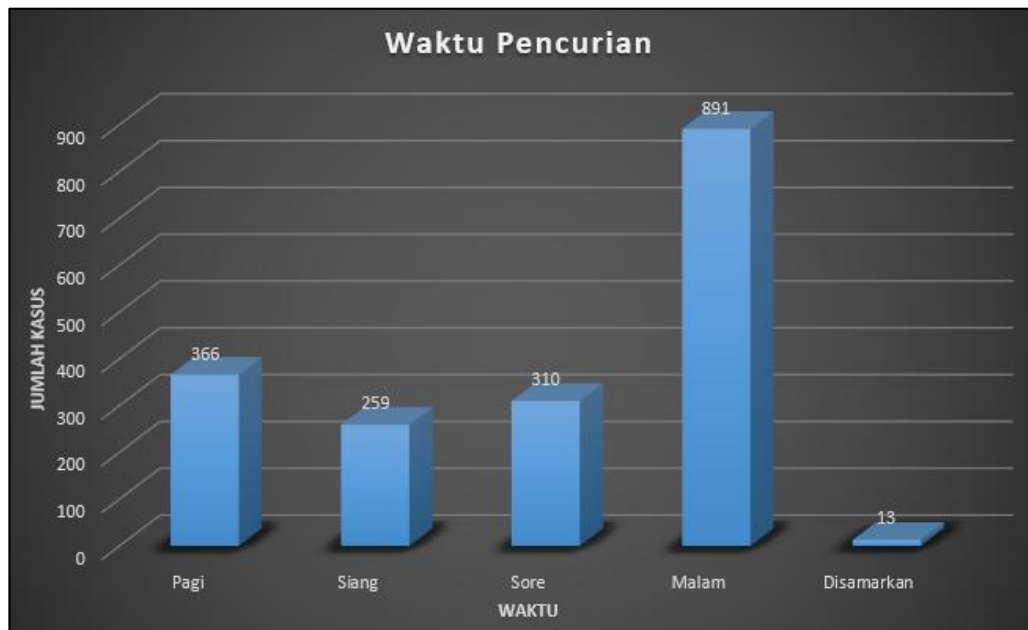
BAB V

HASIL DAN PEMBAHASAN

Pada bab ini, peneliti akan menguraikan hasil penelitian mengenai *Topic Modeling* dengan menggunakan metode *Latent Dirichlet Allocation* (LDA) pada dokumen Surat Dakwaan Kasus Tindak Pidana Pencurian di Pengadilan Negeri Palembang Tahun 2019-2022.

5.1. Analisis Deskriptif

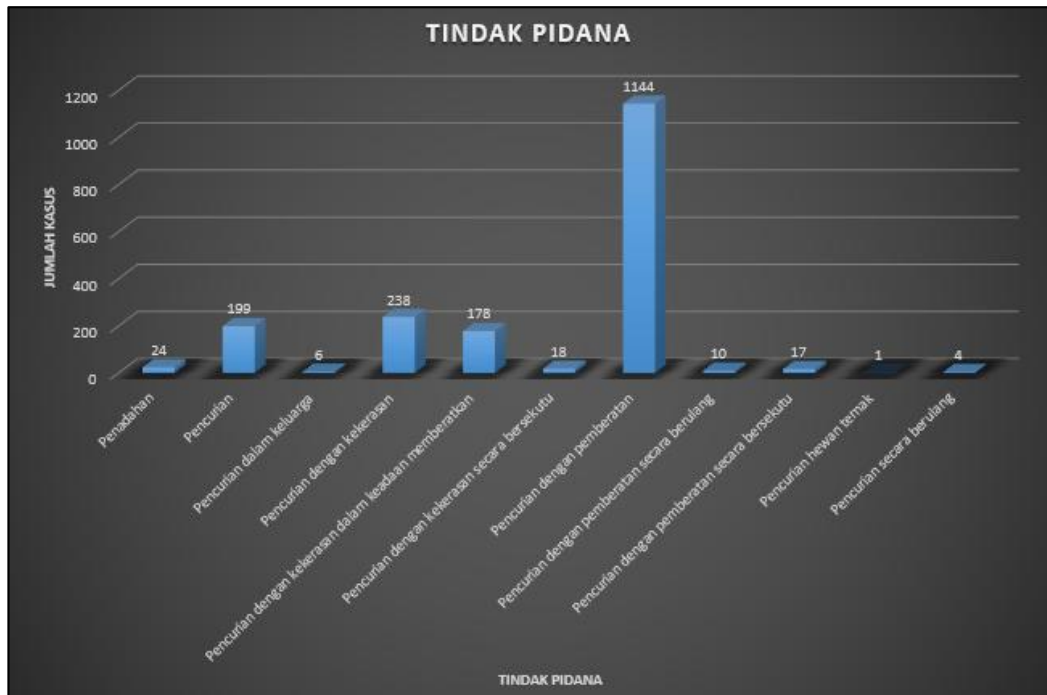
Analisis deskriptif pada penelitian ini adalah untuk mengetahui gambaran umum dari data yang digunakan yaitu pada dokumen Berita Acara Kasus Tindak Pidana Pencurian yang ada di Pengadilan Negeri Palembang tahun 2019 – 2022 yang didapatkan melalui *website* <https://putusan3.mahkamahagung.go.id>.



Gambar 5.1 Waktu Terjadinya Pencurian Tahun 2019 – Tahun 2022

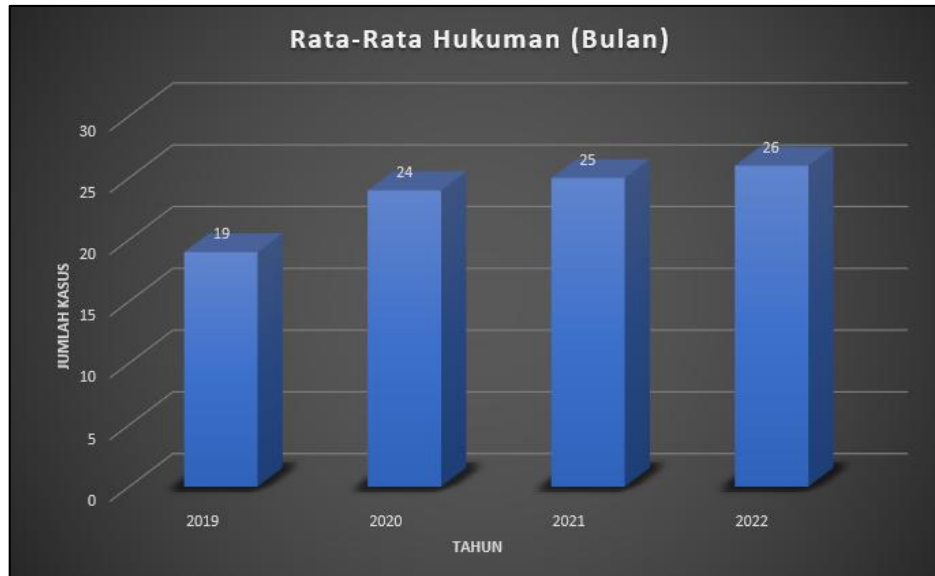
Gambar 5.1 menampilkan waktu terjadinya pencurian tahun 2019 sampai tahun 2022, sepanjang tahun 2019 hingga tahun 2022 dari 1.838 kasus, diketahui bahwa waktu dengan jumlah kasus pencurian paling banyak adalah pada malam hari sedangkan waktu dengan jumlah kasus pencurian paling sedikit adalah pada siang hari. Banyaknya pencurian yang terjadi di malam hari dikarenakan suasana gelap di malam hari membatasi penglihatan dan kemampuan mengawasi sekitar juga berkurang, maka dari itu, para pelaku dapat dengan mudah melakukan tindak

pidana pencurian. Selain itu, malam hari merupakan saat dimana orang-orang beristirahat atau tidur, sehingga, mereka tidak memiliki upaya untuk mencegah tindak pidana pencurian. Pada siang hari, tindak pidana pencurian lebih sedikit terjadi dikarenakan pada waktu tersebut banyak orang-orang beraktivitas dan berlalu-lalang, serta memungkinkan ada saksi mata yang dapat melihat atau mencurigai pelaku, sehingga dapat lebih meminimalisir tindak pidana pencurian.



Gambar 5.2 Tindak Pidana yang Ditetapkan Kepada Terdakwa

Gambar 5.2 menampilkan jenis tindak pidana yang ditetapkan kepada terdakwa. Diketahui, jenis tindak pidana pencurian dengan jumlah kasus tertinggi adalah tindak pidana pencurian dengan pemberatan. Terdapat beberapa faktor yang dapat membuat suatu kasus digolongkan sebagai kasus tindak pidana pencurian dengan pemberatan, salah satunya adalah pencurian yang dilakukan di waktu malam hari. Sehubungan dengan **Gambar 5.1**, dimana kasus pencurian paling banyak terjadi di malam hari, sehingga hal ini juga dapat mempengaruhi tingginya jumlah kasus tindak pidana pencurian dengan pemberatan.



Gambar 5.3 Rata-Rata Hukuman (Bulan) Kasus Pencurian Tahun 2019 – Tahun 2022

Pada **Gambar 5.3** menampilkan informasi mengenai rata-rata hukuman (bulan) kasus pencurian yang terjadi pada tahun 2019 – tahun 2021. Diketahui, tahun 2019 memiliki rata-rata lama hukuman (bulan) yang lebih rendah jika dibandingkan dengan tahun lainnya, sedangkan tahun 2022 memiliki rata-rata lama hukuman (bulan) yang paling tinggi diantara tahun lainnya. Pada tahun 2019 memiliki rata-rata lama hukuman (bulan) yang paling rendah dikarenakan kategori pencurian yang dapat memberatkan hukuman para terdakwa seperti kasus pencurian dengan kekerasan dan lain sebagainya, lebih sedikit terjadi jika dibandingkan dengan ketiga tahun lainnya. Proses selanjutnya akan dilakukan tahapan penyiapan data melalui *preprocessing*.

5.2. Preprocessing Data

Data yang dikumpulkan merupakan data mentah yang tentunya mengandung karakter yang tidak dapat dimengerti komputer. Sehingga diperlukan tahap *preprocessing data*. *Preprocessing data* merupakan tahapan awal yang dilakukan untuk menghapus atau memperbaiki kata agar data lebih terstruktur sehingga dapat dipahami oleh komputer. Adapun tahapan yang dilakukan antara lain *case folding*, *remove punctuation*, *remove number*, *stopword removal*, *stemming*, dan *tokenizing*. Berikut merupakan salah satu dokumen surat dakwaan dari data tahun 2022 yang peneliti pilih sebagai contoh tahap *preprocessing data*.

Tabel 5.1 Data Awal

Surat Dakwaan

Bahwa Ia terdakwa DANDI SAPUTRA BIN RUDI, bersama BOBY (DPO), Pada hari rabu tanggal 15 September 2021 sekira Jam 01.30 WIB atau pada suatu waktu dalam bulan September 2021, di Jalan Faqih Usman Lr. Sei Goren II Rt.28 Rw.05 Kelurahan 2 Ulu Kecamatan Seberang Ulu I Palembang atau disuatu tempat yang masih termasuk dalam daerah hukum Pengadilan Negeri Palembang yang berwenang memeriksa dan mengadili perkara ini, mengambil sesuatu barang berupa 1 Unit Handphone merk VIVO Y 15 warna biru, yang seluruhnya atau sebagian termasuk kepunyaan Saksi korban AHMAD ZAKARIA BIN KM.AKIB atau kepunyaan orang lain selain dari ia terdakwa, dengan maksud akan memiliki barang itu dengan melawan hak, yang didahului, disertai atau diikuti dengan kekerasan, atau ancaman kekerasan terhadap orang dengan maksud untuk mempersiapkan atau mempermudah pencurian itu atau dalam hal tertangkap tangan untuk memungkinkan melarikan diri sendiri atau peserta lainnya, atau untuk tetap menguasai barang yang dicuri, dilakukan oleh dua orang atau lebih dengan bersekutu, Perbuatan mana dilakukan terdakwa dengan cara – cara sebagai berikut : Bahwa pada waktu dan tempat seperti yang diuraikan dalam dakwaan tersebut diatas, berawal terdakwa DANDI SAPUTRA BIN RUDI bersama BOBY (DPO) sedang mengendarai sepeda motor Yamaha NMAX warna hitam milik BOBI (DPO), yang mana BOBI (DPO) yang mengemudikan sepeda motor sedangkan terdakwa dibonceng, sewaktu berada di Jalan Faqih Usman Lr. Sei Goren II Rt.28 Rw.05 Kelurahan 2 Ulu Kecamatan Seberang Ulu I Palembang, melihat korban sedang memainkan handphone didepan rumahnya lalu timbul niat mereka terdakwa untuk mengambil handphone milik korban tersebut, kemudian BOBI (DPO) memutar sepeda motor mendekati korban, Kemudian terdakwa yang masih duduk diatas sepeda motor berpura – pura menanyakan kepada korban dengan berkata “ Kak Numpang Nanya rumah FEBI dimana “ dan dijawab korban tidak tau. Kemudian terdakwa langsung mengambil secara paksa handphone yang korban pegang dengan cara merampas handphone dari tangan korban dan korban pun langsung memeluk terdakwa sambil berteriak “ maling-maling “ dan korban ikut terseret dijalan sehingga sepeda motor yang

digunakan mereka terdakwa menabrak rumah warga lalu terdakwa terjatuh dari sepeda motor dan berhasil ditangkap sedangkan BOBI (DPO) dapat melarikan diri beserta sepeda motor ; Bahwa Akibat perbuatan terdakwa tersebut korban AHMAD ZAKARIA BIN KM.AKIB mengalami kehilangan 1 Unit Handphone merk VIVO Y 15 warna biru yang ditaksir kerugian lebih kurang sebesar Rp.1.500.000.,- (satu juta lima ratus ribu rupiah) ; Bahwa Perbuatan terdakwa sebagaimana diatur dan diancam pidana dalam Pasal 365 ayat (1), Ayat (2) ke 2 kuhp

5.2.1 Case Folding

Case folding merupakan proses mengubah karakter yang semulanya huruf besar menjadi huruf kecil. Tujuan dari tahap ini adalah apabila terdapat kata yang sama, namun dengan cara penulisan yang berbeda antara huruf besar dan huruf kecil maka kata tersebut akan tetap terdeteksi sebagai kata yang sama. Adapun hasil dari *case folding* seperti pada **Tabel 5.2**.

Tabel 5.2 Case Folding

Sebelum
<p>Bahwa Ia terdakwa DANDI SAPUTRA BIN RUDI, bersama BOBY (DPO), Pada hari rabu tanggal 15 September 2021 sekira Jam 01.30 WIB atau pada suatu waktu dalam bulan September 2021, di Jalan Faqih Usman Lr. Sei Goren II Rt.28 Rw.05 Kelurahan 2 Ulu Kecamatan Seberang Ulu I Palembang atau disuatu tempat yang masih termasuk dalam daerah hukum Pengadilan Negeri Palembang yang berwenang memeriksa dan mengadili perkara ini, mengambil sesuatu barang berupa 1 Unit Handphone merk VIVO Y 15 warna biru, yang seluruhnya atau sebagian termasuk kepunyaan Saksi korban AHMAD ZAKARIA BIN KM.AKIB atau kepunyaan orang lain selain dari ia terdakwa, dengan maksud akan memiliki barang itu dengan melawan hak, yang didahului, disertai atau diikuti dengan kekerasan, atau ancaman kekerasan terhadap orang dengan maksud untuk mempersiapkan atau mempermudah pencurian itu atau dalam hal tertangkap tangan untuk memungkinkan melarikan diri sendiri atau peserta lainnya, atau untuk tetap menguasai barang yang dicuri, dilakukan oleh dua orang atau lebih dengan bersekutu, Perbuatan mana dilakukan terdakwa dengan cara -</p>

cara sebagai berikut : Bahwa pada waktu dan tempat seperti yang diuraikan dalam dakwaan tersebut diatas, berawal terdakwa **DANDI SAPUTRA BIN RUDI** bersama **BOBY (DPO)** sedang mengendarai sepeda motor **Yamaha NMAX** warna hitam milik **BOBI (DPO)**, yang mana **BOBI (DPO)** yang mengemudikan sepeda motor sedangkan terdakwa dibonceng, sewaktu berada di **Jalan Faqih Usman Lr. Sei Goren II Rt.28 Rw.05 Kelurahan 2 Ulu Kecamatan Seberang Ulu I Palembang**, melihat korban sedang memainkan handphone didepan rumahnya lalu timbul niat mereka terdakwa untuk mengambil handphone milik korban tersebut, kemudian **BOBI (DPO)** memutar sepeda motor mendekati korban, Kemudian terdakwa yang masih duduk diatas sepeda motor berpura - pura menanyakan kepada korban dengan berkata “**Kak Numpang Nanya rumah FEBI dimana**” dan dijawab korban tidak tau. Kemudian terdakwa langsung mengambil secara paksa handphone yang korban pegang dengan cara merampas handphone dari tangan korban dan korban pun langsung memeluk terdakwa sambil berteriak “**maling-maling**” dan korban ikut terseret dijalan sehingga sepeda motor yang digunakan mereka terdakwa menabrak rumah warga lalu terdakwa terjatuh dari sepeda motor dan berhasil ditangkap sedangkan **BOBI (DPO)** dapat melarikan diri beserta sepeda motor ; Bahwa Akibat perbuatan terdakwa tersebut korban **AHMAD ZAKARIA BIN KM.AKIB** mengalami kehilangan 1 Unit Handphone merk **VIVO Y 15** warna biru yang ditaksir kerugian lebih kurang sebesar **Rp.1.500.000,-** (satu juta lima ratus ribu rupiah) ; Bahwa Perbuatan terdakwa sebagaimana diatur dan diancam pidana dalam **Pasal 365 ayat (1), Ayat (2) ke 2** kuhp

Sesudah

bahwa ia terdakwa **dandi saputra bin rudi**, bersama **boby (dpo)**, pada hari rabu tanggal **15 september 2021** sekira **jam 01.30 wib** atau pada suatu waktu dalam bulan **september 2021**, di **jalan faqih usman lr. sei goren ii rt.28 rw.05 kelurahan 2 ulu kecamatan seberang ulu i palembang** atau disuatu tempat yang masih termasuk dalam daerah hukum **pengadilan negeri palembang** yang berwenang memeriksa dan mengadili perkara ini, mengambil sesuatu barang berupa 1 unit **handphone merk vivo y 15** warna biru, yang seluruhnya atau sebagian termasuk kepunyaan **saksi korban ahmad zakaria bin km.akib** atau kepunyaan orang lain

selain dari ia terdakwa, dengan maksud akan memiliki barang itu dengan melawan hak, yang didahului, disertai atau diikuti dengan kekerasan, atau ancaman kekerasan terhadap orang dengan maksud untuk mempersiapkan atau mempermudah pencurian itu atau dalam hal tertangkap tangan untuk memungkinkan melarikan diri sendiri atau peserta lainnya, atau untuk tetap menguasai barang yang dicuri, dilakukan oleh dua orang atau lebih dengan bersekutu, perbuatan mana dilakukan terdakwa dengan cara - cara sebagai berikut : bahwa pada waktu dan tempat seperti yang diuraikan dalam dakwaan tersebut diatas, berawal terdakwa dandi saputra bin rudi bersama boby (dpo) sedang mengendarai sepeda motor yamaha nmax warna hitam milik bobi (dpo), yang mana bobi (dpo) yang mengemudikan sepeda motor sedangkan terdakwa dibonceng, sewaktu berada di jalan faqih usman Ir. Sei goren ii rt.28 rw.05 kelurahan 2 ulu kecamatan seberang ulu i palembang, melihat korban sedang memainkan handphone didepan rumahnya lalu timbul niat mereka terdakwa untuk mengambil handphone milik korban tersebut, kemudian bobi (dpo) memutar sepeda motor mendekati korban, kemudian terdakwa yang masih duduk diatas sepeda motor berpura - pura menanyakan kepada korban dengan berkata “kak numpang nanya rumah febi dimana” dan dijawab korban tidak tau. kemudian terdakwa langsung mengambil secara paksa handphone yang korban pegang dengan cara merampas handphone dari tangan korban dan korban pun langsung memeluk terdakwa sambil berteriak “maling-maling” dan korban ikut terseret dijalan sehingga sepeda motor yang digunakan mereka terdakwa menabrak rumah warga lalu terdakwa terjatuh dari sepeda motor dan berhasil ditangkap sedangkan bobi (dpo) dapat melarikan diri beserta sepeda motor bahwa akibat perbuatan terdakwa tersebut korban ahmad zakaria bin km.akib mengalami kehilangan 1 unit handphone merk vivo y 15 warna biru yang ditaksir kerugian lebih kurang sebesar rp.1.500.000,- (satu juta lima ratus ribu rupiah) bahwa perbuatan terdakwa sebagaimana diatur dan diancam pidana dalam pasal 365 ayat (1), ayat (2) ke 2 kuhp

5.2.2 Remove Punctuation

Remove Punctuation adalah tahap yang bertujuan untuk menghapus tanda baca yang ada pada dokumen, yaitu titik, koma, tanda seru, semicolon, dan lain sebagainya. Adapun hasil dari *remove punctuation* seperti pada **Tabel 5.3**.

Tabel 5.3 *Remove Punctuation*

Sebelum
bahwa ia terdakwa dandi saputra bin rudi, bersama boby (dpo), pada hari rabu tanggal 15 september 2021 sekira jam 01.30 wib atau pada suatu waktu dalam bulan september 2021, di jalan faqih usman lr. sei goren ii rt.28 rw.05 kelurahan 2 ulu kecamatan seberang ulu i palembang atau disuatu tempat yang masih termasuk dalam daerah hukum pengadilan negeri palembang yang berwenang memeriksa dan mengadili perkara ini, mengambil sesuatu barang berupa 1 unit handphone merk vivo y 15 warna biru, yang seluruhnya atau sebagian termasuk kepunyaan saksi korban ahmad zakaria bin km.akib atau kepunyaan orang lain selain dari ia terdakwa, dengan maksud akan memiliki barang itu dengan melawan hak, yang didahului, disertai atau diikuti dengan kekerasan, atau ancaman kekerasan terhadap orang dengan maksud untuk mempersiapkan atau mempermudah pencurian itu atau dalam hal tertangkap tangan untuk memungkinkan melarikan diri sendiri atau peserta lainnya, atau untuk tetap menguasai barang yang dicuri, dilakukan oleh dua orang atau lebih dengan bersekutu, perbuatan mana dilakukan terdakwa dengan cara-cara sebagai berikut : bahwa pada waktu dan tempat seperti yang diuraikan dalam dakwaan tersebut diatas, berawal terdakwa dandi saputra bin rudi bersama boby (dpo) sedang mengendarai sepeda motor yamaha nmax warna hitam milik bobi (dpo), yang mana bobi (dpo) yang mengemudikan sepeda motor sedangkan terdakwa dibonceng, sewaktu berada di jalan faqih usman lr. sei goren ii rt.28 rw.05 kelurahan 2 ulu kecamatan seberang ulu i palembang, melihat korban sedang memainkan handphone didepan rumahnya lalu timbul niat mereka terdakwa untuk mengambil handphone milik korban tersebut, kemudian bobi (dpo) memutar sepeda motor mendekati korban, kemudian terdakwa yang masih duduk diatas sepeda motor berpura-pura menanyakan kepada korban dengan berkata "kak numpang nanya rumah febi dimana" dan dijawab korban tidak tau.

kemudian terdakwa langsung mengambil secara paksa handphone yang korban pegang dengan cara merampas handphone dari tangan korban dan korban pun langsung memeluk terdakwa sambil berteriak “maling-maling” dan korban ikut terseret dijalan sehingga sepeda motor yang digunakan mereka terdakwa menabrak rumah warga lalu terdakwa terjatuh dari sepeda motor dan berhasil ditangkap sedangkan bobi (dpo) dapat melarikan diri beserta sepeda motor bahwa akibat perbuatan terdakwa tersebut korban ahmad zakaria bin km.akib mengalami kehilangan 1 unit handphone merk vivo y 15 warna biru yang ditaksir kerugian lebih kurang sebesar rp.1.500.000.- (satu juta lima ratus ribu rupiah) bahwa perbuatan terdakwa sebagaimana diatur dan diancam pidana dalam pasal 365 ayat (1), ayat (2) ke 2 kuhp

Sesudah

bahwa ia terdakwa dandi saputra bin rudi bersama boby (dpo) pada hari rabu tanggal 15 september 2021 sekira jam 01:30 wib atau pada suatu waktu dalam bulan september 2021 di jalan faqih usman lr sei goren ii rt.28 rw.05 kelurahan 2 ulu kecamatan seberang ulu i palembang atau disuatu tempat yang masih termasuk dalam daerah hukum pengadilan negeri palembang yang berwenang memeriksa dan mengadili perkara ini mengambil sesuatu barang berupa 1 unit handphone merk vivo y 15 warna biru yang seluruhnya atau sebagian termasuk kepunyaan saksi korban ahmad zakaria bin km.akib atau kepunyaan orang lain selain dari ia terdakwa dengan maksud akan memiliki barang itu dengan melawan hak yang didahului disertai atau diikuti dengan kekerasan atau ancaman kekerasan terhadap orang dengan maksud untuk mempersiapkan atau mempermudah pencurian itu atau dalam hal tertangkap tangan untuk memungkinkan melarikan diri sendiri atau peserta lainnya atau untuk tetap menguasai barang yang dicuri dilakukan oleh dua orang atau lebih dengan bersekutu perbuatan mana dilakukan terdakwa dengan cara-cara sebagai berikut bahwa pada waktu dan tempat seperti yang diuraikan dalam dakwaan tersebut diatas berawal terdakwa dandi saputra bin rudi bersama boby (dpo) sedang mengendarai sepeda motor yamaha nmax warna hitam milik bobi (dpo) yang mana bobi (dpo) yang mengemudikan sepeda motor sedangkan terdakwa dibonceng sewaktu berada di jalan faqih usman lr sei goren ii rt.28 rw.05

kelurahan 2 ulu kecamatan seberang ulu i Palembang melihat korban sedang memainkan handphone didepan rumahnya lalu timbul niat mereka terdakwa untuk mengambil handphone milik korban tersebut kemudian bobi dpo memutar sepeda motor mendekati korban kemudian terdakwa yang masih duduk diatas sepeda motor berpura pura menanyakan kepada korban dengan berkata kak numpang nanya rumah febi dimana dan dijawab korban tidak tau kemudian terdakwa langsung mengambil secara paksa handphone yang korban pegang dengan cara merampas handphone dari tangan korban dan korban pun langsung memeluk terdakwa sambil berteriak maling maling dan korban ikut terseret dijalan sehingga sepeda motor yang digunakan mereka terdakwa menabrak rumah warga lalu terdakwa terjatuh dari sepeda motor dan berhasil ditangkap sedangkan bobi dpo dapat melarikan diri beserta sepeda motor bahwa akibat perbuatan terdakwa tersebut korban ahmad zakaria bin km akib mengalami kehilangan 1 unit handphone merk vivo y 15 warna biru yang ditaksir kerugian lebih kurang sebesar rp 1.500.000 satu jutah lima ratus ribu rupiah bahwa perbuatan terdakwa sebagaimana diatur dan diancam pidana dalam pasal 365 ayat 1 ayat 2 ke 2 kuhp

5.2.3 Remove Number

Remove Number merupakan tahapan *preprocessing* yang dilakukan untuk menghapus angka yang ada pada dokumen surat dakwaan. Adapun hasil dari *remove number* seperti pada **Tabel 5.4**.

Tabel 5.4 *Remove Number*

Sebelum
<p>bahwa ia terdakwa dandi saputra bin rudi bersama boby dpo pada hari rabu tanggal 15 september 2021 sekira jam 01 30 wib atau pada suatu waktu dalam bulan september 2021 di jalan faqih usman lr sei goren ii rt 28 rw 05 kelurahan 2 ulu kecamatan seberang ulu i Palembang atau disuatu tempat yang masih termasuk dalam daerah hukum pengadilan negeri Palembang yang berwenang memeriksa dan mengadili perkara ini mengambil sesuatu barang berupa 1 unit handphone merk vivo y 15 warna biru yang seluruhnya atau sebagian termasuk kepunyaan saksi korban ahmad zakaria bin km akib atau kepunyaan orang lain</p>

selain dari ia terdakwa dengan maksud akan memiliki barang itu dengan melawan hak yang didahului disertai atau diikuti dengan kekerasan atau ancaman kekerasan terhadap orang dengan maksud untuk mempersiapkan atau mempermudah pencurian itu atau dalam hal tertangkap tangan untuk memungkinkan melarikan diri sendiri atau peserta lainnya atau untuk tetap menguasai barang yang dicuri dilakukan oleh dua orang atau lebih dengan bersekutu perbuatan mana dilakukan terdakwa dengan cara-cara sebagai berikut bahwa pada waktu dan tempat seperti yang diuraikan dalam dakwaan tersebut diatas berawal terdakwa dandi saputra bin rudi bersama boby dpo sedang mengendarai sepeda motor yamaha nmax warna hitam milik bobi dpo yang mana bobi dpo yang mengemudikan sepeda motor sedangkan terdakwa dibonceng sewaktu berada di jalan faqih usman lr sei goren ii rt 28 rw 05 kelurahan 2 ulu kecamatan seberang ulu i Palembang melihat korban sedang memainkan handphone didepan rumahnya lalu timbul niat mereka terdakwa untuk mengambil handphone milik korban tersebut kemudian bobi dpo memutar sepeda motor mendekati korban kemudian terdakwa yang masih duduk diatas sepeda motor berpura-pura menanyakan kepada korban dengan berkata kak numpang nanya rumah febi diman dan dijawab korban tidak tau kemudian terdakwa langsung mengambil secara paksa handphone yang korban pegang dengan cara merampas handphone dari tangan korban dan korban pun langsung memeluk terdakwa sambil berteriak maling maling dan korban ikut terseret dijalan sehingga sepeda motor yang digunakan mereka terdakwa menabrak rumah warga lalu terdakwa terjatuh dari sepeda motor dan berhasil ditangkap sedangkan bobi dpo dapat melarikan diri beserta sepeda motor bahwa akibat perbuatan terdakwa tersebut korban ahmad zakaria bin km akib mengalami kehilangan 1 unit handphone merk vivo y 15 warna biru yang ditaksir kerugian lebih kurang sebesar rp 1 500 000 satu jutah lima ratus ribu rupiah bahwa perbuatan terdakwa sebagaimana diatur dan diancam pidana dalam pasal 365 ayat 1 ayat 2 ke 2 kuhp

Sesudah

bahwa ia terdakwa dandi saputra bin rudi bersama boby dpo pada hari rabu tanggal [] september [] sekira jam [] [] wib atau pada suatu waktu dalam bulan september [] di jalan faqih usman lr sei goren ii rt [] rw [] kelurahan [] ulu kecamatan seberang ulu i Palembang atau disuatu tempat yang masih termasuk dalam daerah hukum pengadilan negeri Palembang yang berwenang memeriksa dan mengadili perkara ini mengambil sesuatu barang berupa [] unit handphone merk vivo y [] warna biru yang seluruhnya atau sebagian termasuk kepunyaan saksi korban ahmad zakaria bin km akib atau kepunyaan orang lain selain dari ia terdakwa dengan maksud akan memiliki barang itu dengan melawan hak yang didahului disertai atau diikuti dengan kekerasan atau ancaman kekerasan terhadap orang dengan maksud untuk mempersiapkan atau mempermudah pencurian itu atau dalam hal tertangkap tangan untuk memungkinkan melarikan diri sendiri atau peserta lainnya atau untuk tetap menguasai barang yang dicuri dilakukan oleh dua orang atau lebih dengan bersekutu perbuatan mana dilakukan terdakwa dengan cara [] cara sebagai berikut bahwa pada waktu dan tempat seperti yang diuraikan dalam dakwaan tersebut diatas berawal terdakwa dandi saputra bin rudi bersama boby dpo sedang mengendarai sepeda motor yamaha nmax warna hitam milik bobi dpo yang mana bobi dpo yang mengemudikan sepeda motor sedangkan terdakwa dibonceng sewaktu berada di jalan faqih usman lr sei goren ii rt [] rw [] kelurahan [] ulu kecamatan seberang ulu i Palembang melihat korban sedang memainkan handphone didepan rumahnya lalu timbul niat mereka terdakwa untuk mengambil handphone milik korban tersebut kemudian bobi dpo memutar sepeda motor mendekati korban kemudian terdakwa yang masih duduk diatas sepeda motor berpura-pura menanyakan kepada korban dengan berkata kak numpang nanya rumah febi dimana dan dijawab korban tidak tau kemudian terdakwa langsung mengambil secara paksa handphone yang korban pegang dengan cara merampas handphone dari tangan korban dan korban pun langsung memeluk terdakwa sambil berteriak maling maling dan korban ikut terseret dijalan sehingga sepeda motor yang digunakan mereka terdakwa menabrak rumah warga lalu terdakwa terjatuh dari sepeda motor dan berhasil ditangkap sedangkan bobi dpo dapat melarikan diri beserta sepeda motor bahwa akibat perbuatan terdakwa tersebut korban ahmad zakaria bin km akib

mengalami kehilangan unit handphone merk vivo y warna biru yang ditaksir kerugian lebih kurang sebesar rp satu jutah lima ratus ribu rupiah bahwa perbuatan terdakwa sebagaimana diatur dan diancam pidana dalam pasal ayat ke kuhp

5.2.4 Stopword Removal

Stopword Removal adalah tahapan *preprocessing* yang berfungsi untuk menghapus kata-kata yang sering muncul dalam dokumen surat dakwaan, namun tidak memiliki makna. Contoh kata yang termasuk dalam *stopword* adalah “di”, “yang”, “atau”, dan lain sebagainya. Adapun hasil dari *stopword removal* seperti pada **Tabel 5.5**.

Tabel 5.5 *Stopword Removal*

Sebelum
<p>bahwa ia terdakwa dandi saputra bin rudi bersama boby dpo pada hari rabu tanggal september sekira jam wib atau pada suatu waktu dalam bulan september di jalan faqih usman lr sei goren ii rt rw kelurahan ulu kecamatan seberang ulu i palembang atau disuatu tempat yang masih termasuk dalam daerah hukum pengadilan negeri palembang yang berwenang memeriksa dan mengadili perkara ini mengambil sesuatu barang berupa unit handphone merk vivo y warna biru yang seluruhnya atau sebagian termasuk kepunyaan saksi korban ahmad zakaria bin km akib atau kepunyaan orang lain selain dari ia terdakwa dengan maksud akan memiliki barang itu dengan melawan hak yang didahului disertai atau diikuti dengan kekerasan atau ancaman kekerasan terhadap orang dengan maksud untuk mempersiapkan atau mempermudah pencurian itu atau dalam hal tertangkap tangan untuk memungkinkan melarikan diri sendiri atau peserta lainnya atau untuk tetap menguasai barang yang dicuri dilakukan oleh dua orang atau lebih dengan bersekutu perbuatan mana dilakukan terdakwa dengan cara cara sebagai berikut bahwa pada waktu dan tempat seperti yang diuraikan dalam dakwaan tersebut diatas berawal terdakwa dandi saputra bin rudi bersama boby dpo sedang mengendarai sepeda motor yamaha nmax warna hitam milik bobi dpo yang mana bobi dpo yang mengemudikan sepeda motor sedangkan terdakwa dibonceng sewaktu berada di jalan faqih</p>

usman lr sei goren ii rt rw kelurahan ulu kecamatan seberang ulu i Palembang melihat korban sedang memainkan handphone didepan rumahnya lalu timbul niat mereka terdakwa untuk mengambil handphone milik korban tersebut kemudian bobi dpo memutar sepeda motor mendekati korban kemudian terdakwa yang masih duduk diatas sepeda motor berpura pura menanyakan kepada korban dengan berkata kak numpang nanya rumah febi dimana dan dijawab korban tidak tau kemudian terdakwa langsung mengambil secara paksa handphone yang korban pegang dengan cara merampas handphone dari tangan korban dan korban pun langsung memeluk terdakwa sambil berteriak maling maling dan korban ikut terseret dijalan sehingga sepeda motor yang digunakan mereka terdakwa menabrak rumah warga lalu terdakwa terjatuh dari sepeda motor dan berhasil ditangkap sedangkan bobi dpo dapat melarikan diri beserta sepeda motor bahwa akibat perbuatan terdakwa tersebut korban ahmad zakaria bin km akib mengalami kehilangan unit handphone merk vivo y warna biru yang ditaksir kerugian lebih kurang sebesar rp satu jutah lima ratus ribu rupiah bahwa perbuatan terdakwa sebagaimana diatur dan diancam pidana dalam pasal ayat ayat ke kuhp

Sesudah

jalan sei goren ulu ulu disuatu pengadilan berwenang handphone kepunyaan korban kepunyaan melawan didahului disertai diikuti kekerasan ancaman kekerasan mempermudah tertangkap peserta menguasai bersekutu perbuatan diuraikan dakwaan diatas mengendarai sepeda motor nmax mengemudikan sepeda motor dibonceng jalan sei goren ulu ulu korban memainkan handphone didepan rumahnya handphone korban memutar sepeda motor mendekati korban diatas sepeda motor berpura pura korban numpang nanya korban tau langsung paksa handphone korban merampas handphone korban korban langsung memeluk korban terseret dijalan sepeda motor menabrak terjatuh sepeda motor berhasil beserta sepeda motor akibat perbuatan korban kehilangan handphone kerugian perbuatan diatur pidana pasal ayat ayat kuhp

5.2.5 Stemming

Stemming adalah tahapan untuk mendapatkan kata dasar dari suatu kata, yaitu dengan menghilangkan imbuhan pada kata tersebut. Pada tahapan mengubah kata dasar ini peneliti menggunakan *stemming* dan bukan *lemmatization*, hal tersebut dikarenakan teks yang digunakan merupakan teks berbahasa Indonesia. Bahasa Indonesia memiliki banyak bentuk kata dan imbuhan, oleh karena itu peneliti memilih untuk menggunakan *stemming*. Adapun hasil dari *stemming* seperti pada **Tabel 5.6**.

Tabel 5.6 *Stemming*

Sebelum
<p>jalan sei goren ulu ulu disuatu pengadilan berwenang handphone kepunyaan korban kepunyaan melawan didahului disertai diikuti kekerasan ancaman kekerasan mempermudah tertangkap peserta menguasai bersekutu perbuatan diuraikan dakwaan diatas mengendarai sepeda motor nmax mengemudikan sepeda motor dibonceng jalan sei goren ulu ulu korban memainkan handphone didepan rumahnya handphone korban memutar sepeda motor mendekati korban diatas sepeda motor berpura pura korban numpang nanya korban tau langsung paksa handphone korban merampas handphone korban korban langsung memeluk korban terseret dijalan sepeda motor menabrak terjatuh sepeda motor berhasil beserta sepeda motor akibat perbuatan korban kehilangan handphone kerugian perbuatan diatur pidana pasal ayat ayat kuhp</p>
Sesudah
<p>jalan sei goren ulu ulu suatu adil wenang handphone punya korban punya lawan dahulu serta ikut keras ancam keras mudah tangkap serta kuasa sekutu buat urai dakwa atas kendaraan sepeda motor nmax kemudi sepeda motor bonceng jalan sei goren ulu ulu korban main handphone depan rumah handphone korban putar sepeda motor dekat korban atas sepeda motor pura pura korban numpang nanya korban tau langsung paksa handphone korban rampas handphone korban korban langsung peluk korban seret jalan sepeda motor tabrak jatuh sepeda motor hasil serta sepeda motor akibat buat korban hilang handphone rugi buat atur pidana pasal ayat ayat kuhp</p>

5.2.6 Tokenizing

Tokenizing adalah proses pemisahan kata pada suatu dokumen sehingga memudahkan tahap-tahap yang akan dilakukan selanjutnya. Adapun hasil dari *tokenizing* seperti pada **Tabel 5.7**.

Tabel 5.7 *Tokenizing*

Sebelum
jalan sei goren ulu ulu suatu adil wenang handphone punya korban punya lawan dahulu serta ikut kerasancam keras mudah tangkap serta kuasa sekutu buat urai dakwa atas kendara sepeda motor nmax kemudi sepeda motor bonceng jalan sei goren ulu ulu korban main handphone depan rumah handphone korban putar sepeda motor dekat korban atas sepeda motor pura pura korban numpang nanya korban tau langsung paksa handphone korban rampas handphone korban korban langsung peluk korban seret jalan sepeda motor tabrak jatuh sepeda motor hasil serta sepeda motor akibat buat korban hilang handphone rugi buat atur pidana pasal ayat ayat kuhp
Sesudah
['jalan', 'sei', 'goren', 'ulu', 'ulu', 'suatu', 'adil', 'wenang', 'handphone', 'punya', 'korban', 'punya', 'lawan', 'dahulu', 'serta', 'ikut', 'keras', 'ancam', 'keras', 'mudah', 'tangkap', 'serta', 'kuasa', 'sekutu', 'buat', 'urai', 'dakwa', 'atas', 'kendara', 'sepeda', 'motor', 'nmax', 'kemudi', 'sepeda', 'motor', 'bonceng', 'jalan', 'sei', 'goren', 'ulu', 'ulu', 'korban', 'main', 'handphone', 'depan', 'rumah', 'handphone', 'korban', 'putar', 'sepeda', 'motor', 'dekat', 'korban', 'atas', 'sepeda', 'motor', 'pura', 'pura', 'korban', 'numpang', 'nanya', 'korban', 'tau', 'langsung', 'paksa', 'handphone', 'korban', 'rampas', 'handphone', 'korban', 'korban', 'langsung', 'peluk', 'korban', 'seret', 'jalan', 'sepeda', 'motor', 'tabrak', 'jatuh', 'sepeda', 'motor', 'hasil', 'serta', 'sepeda', 'motor', 'akibat', 'buat', 'korban', 'hilang', 'handphone', 'rugi', 'buat', 'atur', 'pidana', 'pasal', 'ayat', 'ayat', 'kuhp']

5.3. Pembobotan Kata TF-IDF

Pada analisis *Topic Modeling*, data yang digunakan harus berskala numerik, sehingga perlu dilakukan pembobotan kata agar data yang berbentuk teks dapat

menjadi numerik. Tahap pembobotan kata merupakan tahap yang dilakukan setelah proses *preprocessing data*, tahap ini bertujuan untuk mengubah suatu data text menjadi *numeric* dengan menggunakan metode pembobotan TF-IDF (*Term Frequency – Inverse Document Frequency*). Dengan dilakukannya pembobotan data, dapat diketahui pula kata mana yang memiliki frekuensi terbanyak pada suatu dokumen. Dalam penelitian ini, peneliti mengambil sampel kata yang digunakan untuk perhitungan pembobotan kata, yaitu kata Dompot, Laptop, Mobil, Senjata, dan Uang, serta hasil yang ditampilkan merupakan hasil pembobotan kata untuk tahun 2022. Berikut ini merupakan hasil pembobotan kata menggunakan TF-IDF yang didapatkan.

Tabel 5.8 Sampel Hasil Pembobotan Kata TF-IDF Tahun 2022

D	...	Dompot	...	Laptop	...	Mobil	...	Senjata	...	Uang	...
1	...	0	...	0	...	0	...	0	...	3.554	...
2	...	0	...	0	...	0	...	9.946	...	1.777	...
3	...	3.289	...	0	...	0	...	0	...	0	...
⋮	...	⋮	...	⋮	...	⋮	...	⋮	...	⋮	...
135	...	0	...	0	...	0	...	3.315	...	8.885	...
136	...	0	...	27.903	...	33.675	...	0	...	0	...
137	...	0	...	32.554	...	64.289	...	0	...	0	...
138	...	0	...	0	...	0	...	0	...	0	...
139	...	0	...	9.301	...	39.798	...	0	...	0	...
140	...	0	...	0	...	0	...	0	...	0	...
141	...	0	...	0	...	0	...	0	...	1.777	...
⋮	...	⋮	...	⋮	...	⋮	...	⋮	...	⋮	...
193	...	0	...	0	...	21.429	...	0	...	0	...
⋮	...	⋮	...	⋮	...	⋮	...	⋮	...	⋮	...
385	...	0	...	0	...	0	...	0	...	0	...

Tabel 5.8 merupakan *output* TF-IDF yang didapatkan dari hasil perhitungan menggunakan *software Python*. Diketahui, terdapat jumlah dokumen sebanyak 385 dokumen dan jumlah kata sebanyak 5571 kata. Kemudian, peneliti akan menghitung TF-IDF dengan cara manual, dengan menghitung *Term Frequency*

(TF) terlebih dahulu, yaitu menghitung jumlah munculnya suatu kata pada sebuah dokumen.

Tabel 5.9 Hasil Perhitungan Nilai TF

D	...	Dompet	...	Laptop	...	Mobil	...	Senjata	...	Uang	...
1	...	0	...	0	...	0	...	0	...	2	...
2	...	0	...	0	...	0	...	3	...	1	...
3	...	1	...	0	...	0	...	0	...	0	...
⋮	...	⋮	...	⋮	...	⋮	...	⋮	...	⋮	...
135	...	0	...	0	...	0	...	1	...	5	...
136	...	0	...	6	...	11	...	0	...	0	...
137	...	0	...	7	...	21	...	0	...	0	...
138	...	0	...	0	...	0	...	0	...	1	...
139	...	0	...	2	...	13	...	0	...	0	...
140	...	0	...	0	...	0	...	0	...	0	...
141	...	0	...	0	...	0	...	0	...	1	...
⋮	...	⋮	...	⋮	...	⋮	...	⋮	...	⋮	...
193	...	0	...	0	...	7	...	0	...	0	...
⋮	...	⋮	...	⋮	...	⋮	...	⋮	...	⋮	...
385	...	0	...	0	...	0	...	0	...	0	...

Berdasarkan *output* di atas, diketahui bahwa nilai 0 pada **Tabel 5.9** menunjukkan bahwa dalam dokumen tersebut tidak terdapat kata yang dicari. Sedangkan, nilai 1, 2, 3, dan seterusnya, menunjukkan bahwa pada dokumen terdapat kata yang dicari sejumlah tersebut.

Kemudian, untuk perhitungan manual, peneliti akan menggunakan kata “Mobil” pada dokumen ke 193 sebagai contoh kata yang akan dilakukan perhitungan TF-IDF manual. Pada **Tabel 5.9** diketahui bahwa *Term Frequency* dari kata “Mobil” adalah sebanyak 7. Nilai *Document Frequency* untuk kata “Mobil” sebanyak 49, hal ini menunjukkan bahwa kata “Mobil” terdapat pada 49 dokumen. Setelah itu, dilakukan perhitungan untuk nilai IDF (*Inverse Document Frequency*), kata yang memiliki DF yang tinggi akan didapatkan nilai IDF yang rendah. Hal ini bersinggungan dengan perhitungan nilai IDF yang menggunakan rumus seperti pada **Persamaan 3.3**. Dalam perhitungan TF-IDF ini digunakan sampel yaitu TF-

IDF untuk data tahun 2022 dengan jumlah sebanyak 385 dokumen. Berikut merupakan perhitungan nilai IDF untuk kata “Mobil”.

$$IDF_{(mobil)} = \ln\left(\frac{D}{DF_{(mobil)}}\right)$$

$$IDF_{(mobil)} = \ln\left(\frac{385}{49}\right) = 2.061423$$

Dengan diketahuinya nilai TF, DF, dan IDF dari kata “Mobil, selanjutnya peneliti akan menghitung secara manual nilai pembobotan TF-IDF dengan menggunakan rumus seperti pada **Persamaan 3.4**. Berikut merupakan hasil pembobotan TF-IDF dari kata “Mobil” yang didapatkan.

$$w_{(mobil,193)} = TF_{mobil,193} \times (IDF_{mobil} + 1)$$

$$w_{(mobil,193)} = 7 \times (2.061423 + 1) = 21.42996$$

Berdasarkan hasil perhitungan manual, didapatkan bahwa kata “Mobil” pada dokumen ke 193 memiliki pembobotan kata TF-IDF sebesar 21,42996. Hasil perhitungan dapat dirangkum seperti pada **Tabel 5.10**

Tabel 5.10 Hasil Pembobotan Kata Mobil

Term(t)	TF	D	DF _{mobil}	IDF	TF-IDF
	D ₁₉₃				D ₁₉₃
Mobil	7	385	49	2.061423	21.42996

Dengan didapatkannya hasil perhitungan manual, maka diketahui bahwa hasil tersebut sama dengan hasil yang didapatkan dengan perhitungan otomatis menggunakan *software Python*. Sehingga, dapat dilakukan tahapan selanjutnya yaitu *Topic Modeling* menggunakan metode *Latent Dirichlet Allocation (LDA)*.

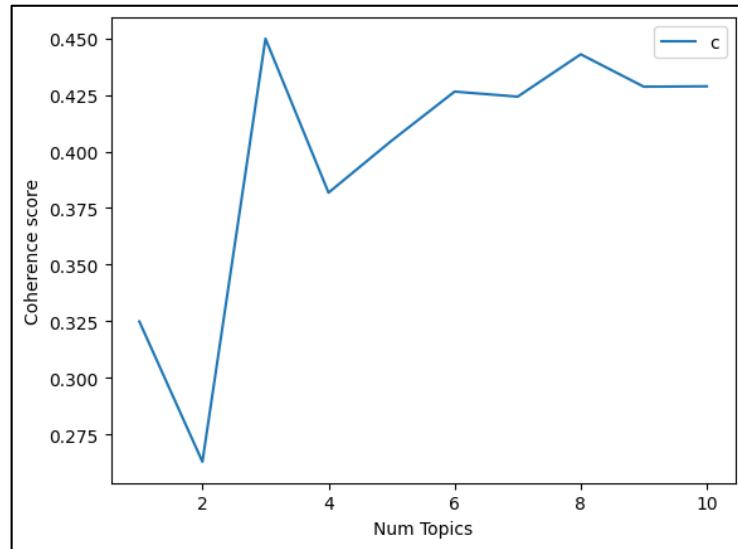
5.4. Topic Modeling dengan Latent Dirichlet Allocation

Topic Modeling dengan metode *Latent Dirichlet Allocation* merupakan salah satu teknik yang digunakan untuk mengidentifikasi topik dalam sebuah kumpulan dokumen. Pada subbab ini akan membahas hasil yang didapatkan dari analisis *Topic Modeling* yang telah peneliti lakukan.

5.4.1 Tahun 2019

Untuk menentukan jumlah topik dalam analisis *Topic Modeling* dapat dilakukan dengan melihat visualisasi pada grafik *coherence score*. *Coherence score* adalah nilai yang digunakan untuk mengukur seberapa baik pemodelan topik,

semakin tinggi nilai *coherence* maka akan semakin baik pula model yang akan didapatkan. Sehingga, yang akan digunakan dalam analisis *Topic Modeling* adalah jumlah topik dengan nilai *coherence* tertinggi. Pada penelitian ini, limit nilai *coherence* yang peneliti gunakan berjumlah 11. Berikut merupakan grafik nilai *coherence* yang didapatkan untuk pemodelan topik tahun 2019.



Gambar 5.4 Grafik Nilai *Coherence* Tahun 2019

Berdasarkan grafik, diketahui bahwa topik dengan *coherence* tertinggi merupakan *num topics* ke-4. **Gambar 5.4** hanya menampilkan grafik saja, sehingga masih belum diketahui dengan pasti berapa nilai *coherence* yang didapatkan pada jumlah topik tersebut. Sehingga, nilai *coherence* untuk masing-masing jumlah topik di atas dapat dilihat pada **Tabel 5.11**

Tabel 5.11 Hasil Nilai *Coherence* Tahun 2019

Num Topics	Topic Coherence	Num Topics	Topic Coherence
1	0.32491	6	0.42644
2	0.26297	7	0.42421
3	0.44983	8	0.44291
4	0.38181	9	0.4286
5	0.40475	10	0.42878

Hasil *topic coherence* pada **Tabel 5.11** menunjukkan bahwa *topic coherence* tertinggi adalah *num topics* ke-3 dengan nilai *coherence* sebesar 0.44983. Berdasarkan nilai *coherence* tertinggi, maka jumlah topik tersebutlah yang akan

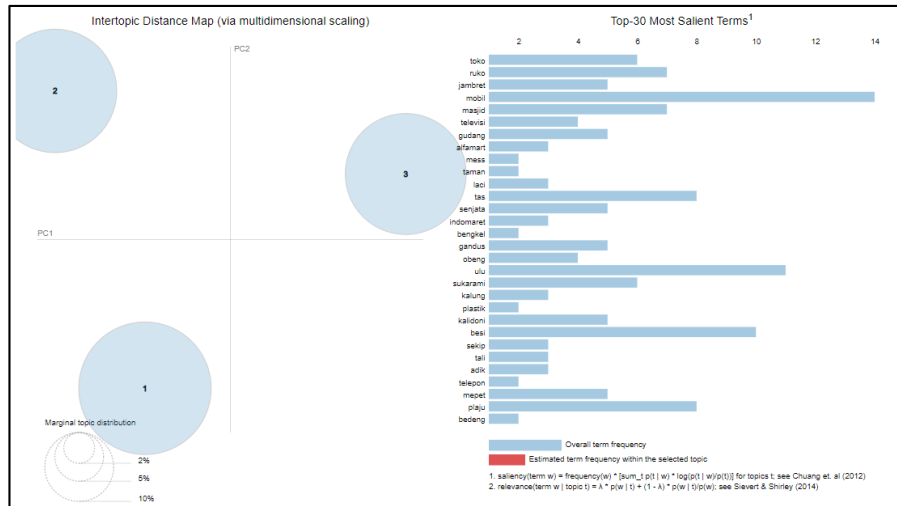
peneliti gunakan sebagai acuan untuk melakukan analisis *Topic Modeling* Kasus Tindak Pidana Pencurian di Pengadilan Negeri Palembang tahun 2019.

Kemudian, letak dari tiap topik pada titik kordinat dapat dilihat berdasarkan nilai *principal component* (PC) yang terbentuk dari sekian banyaknya frekuensi dokumen yang kemudian direduksi dimensinya dengan membentuk variabel-variabel baru (dalam penelitian ini berupa PC1 dan PC2) serta hanya akan direpresentasikan dalam bentuk diagram kartesius. Masing-masing topik bisa berada di kuadran I, kuadran II, kuadran III, atau kuadran IV tergantung dengan nilai PC yang didapatkan. Kuadran pada visualisasi memiliki fungsi untuk memetakan topik, sehingga akan terlihat sejauh mana topik-topik tersebut saling berdekatan atau terpisah, serta untuk mengetahui pola hubungan dari topik-topik tertentu. Topik yang berdekatan dalam kuadran kemungkinan besar memiliki kemiripan dalam hal kata yang muncul atau topik tersebut berkaitan satu sama lainnya.

Tabel 5.12 Hasil Nilai PC Tahun 2019

Topic	PC1	PC2
1	-0.00952	-0.028496
2	-0.02455	0.021190
3	0.03407	0.007307

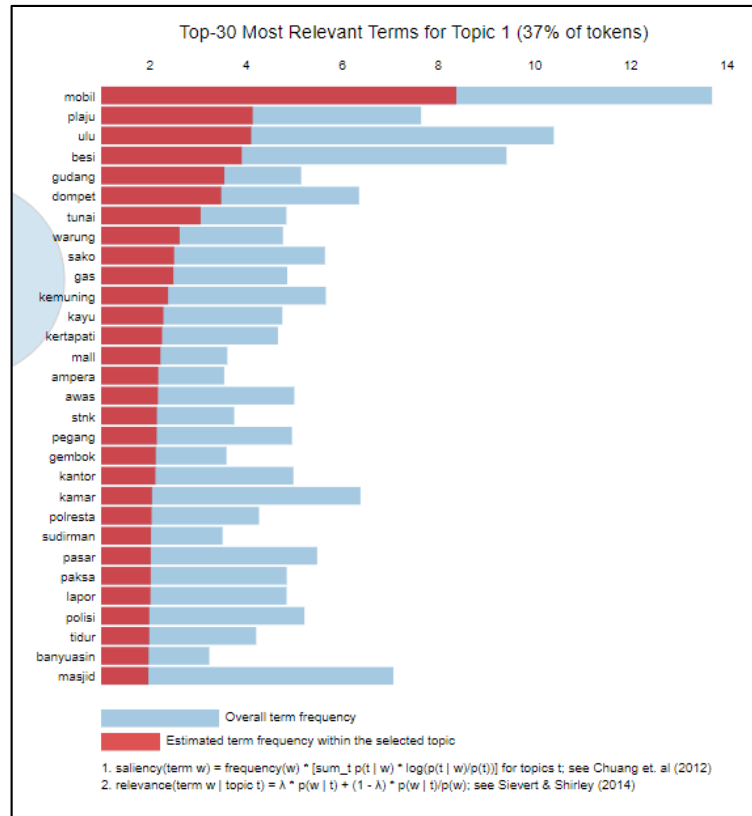
Tabel 5.12 menunjukkan nilai *principal component* (PC) untuk masing-masing topik yang akan digunakan dalam visualisasi *pyLDAvis*. PC1 menunjukkan titik koordinat pada sumbu X, sedangkan nilai PC2 digunakan sebagai titik koordinat pada sumbu Y. Pada topik 1, didapatkan PC1 bernilai negatif dan PC2 bernilai negatif, sehingga topik 1 akan terletak di kuadran 3. Pada topik 2, didapatkan PC1 bernilai negatif dan PC2 bernilai positif, sehingga topik 2 akan terletak di kuadran 2. Pada topik 3, didapatkan PC1 bernilai positif dan PC2 bernilai positif, sehingga topik 3 akan terletak di kuadran 1. Dari **Gambar 5.5**, terlihat bahwa letak dari masing-masing topik sudah sesuai dengan pemosisian berdasarkan nilai PC1 dan PC2 yang telah dijelaskan sebelumnya. Hasil visualisasi dari *pyLDAvis* terlihat seperti berikut.



Gambar 5.5 Visualisasi Ketiga Topik Tahun 2019

Berdasarkan **Gambar 5.5** pada bagian kiri terlihat bahwa letak topik 1, topik 2, dan topik 3 menyebar pada kuadran yang berbeda-beda, sehingga dapat dikatakan ketiga topik tersebut memiliki hubungan antar topik yang lemah. Visualisasi di atas juga menunjukkan *bar chart top 30 global term* paling selaras dengan suatu topik tertentu. *Bar chart* berwarna biru menunjukkan keseluruhan *term frequency* dalam dokumen, sementara itu, *bar chart* berwarna merah menunjukkan *term frequency* dalam suatu topik tertentu.

Analisis pemodelan topik bertujuan untuk melihat topik-topik yang terbentuk dari sekumpulan data teks, dengan melihat *output* atau visualisasi yang dihasilkan, lalu menyimpulkan menjadi suatu topik tertentu. Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-1 yang didapatkan:



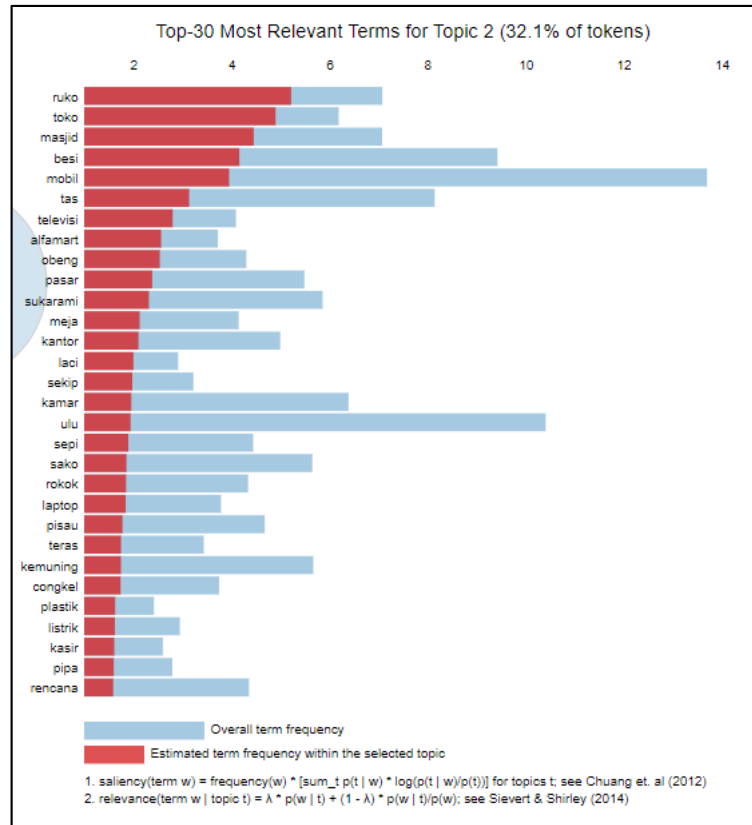
Gambar 5.6 Visualisasi Topik 1 Tahun 2019

Pada **Gambar 5.6**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 1, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 1 adalah sebesar 37%, yang artinya pentingnya topik terhadap dokumen sebesar 37%.

Tabel 5.13 Model Topik 1 Tahun 2019

Model Topik 1
$ \begin{aligned} &0.020 * "mobil" + 0.010 * "plaju" + 0.010 * "ulu" + 0.009 * "besi" + \\ &0.008 * "gudang" + 0.008 * "dompet" + 0.007 * "tunai" + 0.006 * "warung" \\ &+ 0.006 * "sako" + 0.007 * "gas" \end{aligned} $

Tabel 5.13 menampilkan model pada topik 1, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 1 dari analisis LDA tahun 2019 membahas mengenai jenis barang curian. Hal ini didasari kata ‘dompet’, ‘tunai’, dan ‘gas’. Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, pada kasus ini terdakwa mengambil tas milik korban yang berisikan **dompet** yang berisikan uang **tunai** sebesar Rp.7,000,- (tujuh ribu rupiah), alat kosmetik, dan 1 (satu) unit *handphone*.



Gambar 5.7 Visualisasi Topik 2 Tahun 2019

Pada **Gambar 5.7**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 2, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 2 adalah sebesar 32.1%, yang artinya pentingnya topik terhadap dokumen sebesar 32.1%.

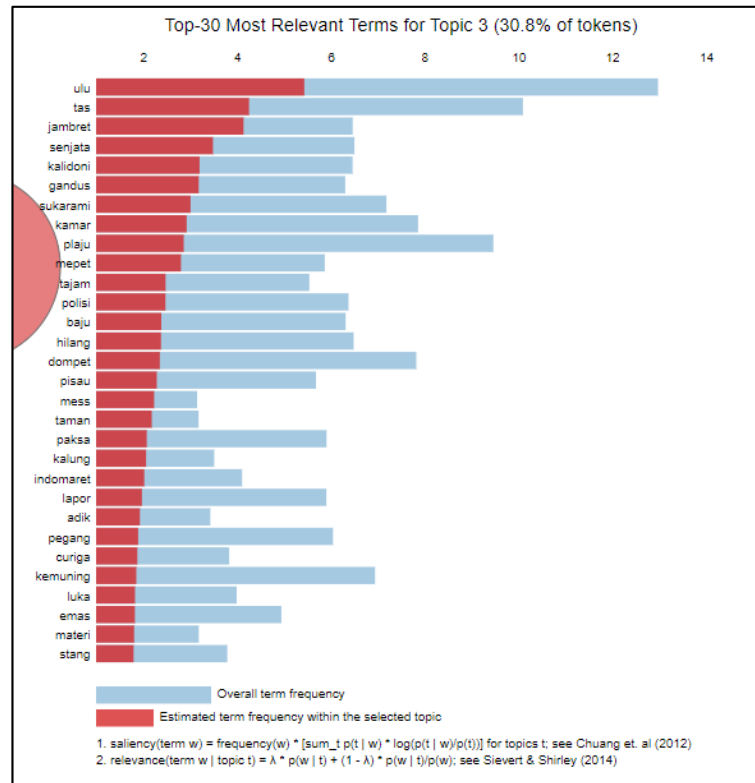
Tabel 5.14 Model Topik 2 Tahun 2019

Model Topik 2
$0.014 * \text{"ruko"} + 0.013 * \text{"toko"} + 0.012 * \text{"masjid"} + 0.011 * \text{"besi"} +$ $0.011 * \text{"mobil"} + 0.008 * \text{"tas"} + 0.007 * \text{"televisi"} +$ $0.007 * \text{"alfamart"} + 0.007 * \text{"obeng"} + 0.006 * \text{"pasar"}$

Tabel 5.14 menampilkan model pada topik 2, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 2 dari analisis LDA tahun 2019 membahas mengenai alat pendukung pencurian. Hal ini didasari kata ‘besi’ dan ‘obeng’ yang mengindikasikan bahwa kedua alat tersebut digunakan pelaku untuk mempermudah melaksanakan aksinya. Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, pada suatu kasus terdakwa menggunakan **obeng** untuk merusak jendela rumah korban, pada kasus lainnya

terdakwa menggunakan **obeng** untuk mempermudah dalam membuka bracket TV sehingga terdakwa dapat membawa barang curian tersebut.

Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-3 yang didapatkan:



Gambar 5.8 Visualisasi Topik 3 Tahun 2019

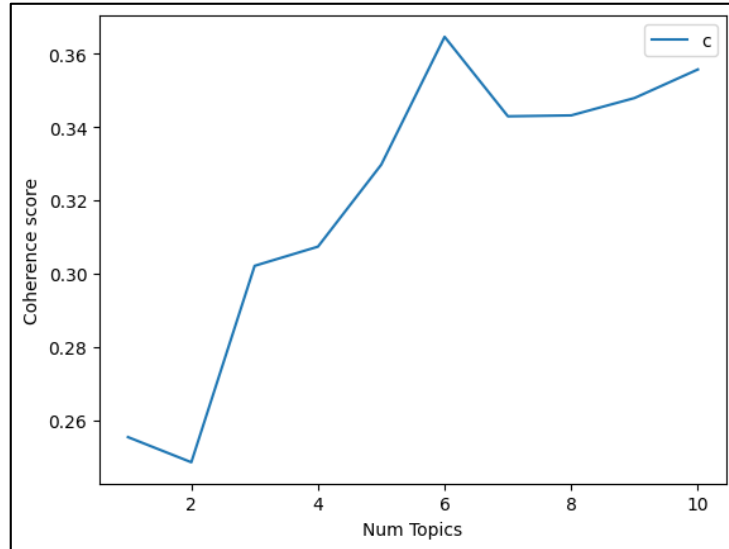
Pada **Gambar 5.8**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 3, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 3 adalah sebesar 30.8%, yang artinya pentingnya topik terhadap dokumen sebesar 30.8%.

Tabel 5.15 Model Topik 3 Tahun 2019

Model Topik 3
$0.012 * \text{"ulu"} + 0.010 * \text{"tas"} + 0.010 * \text{"jambret"} + 0.008 * \text{"senjata"} + 0.008 * \text{"kalidoni"} + 0.007 * \text{"gandum"} + 0.007 * \text{"sukarami"} + 0.007 * \text{"kamar"} + 0.007 * \text{"plaju"} + 0.007 * \text{"mepet"}$

Tabel 5.15 menampilkan model pada topik 3, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 3 dari analisis LDA tahun 2019 membahas mengenai metode pencurian. Hal ini didasari kata ‘jambret’ dan

semakin tinggi nilai *coherence* maka akan semakin baik pula model yang akan didapatkan. Sehingga, yang akan digunakan dalam analisis *Topic Modeling* adalah jumlah topik dengan nilai *coherence* tertinggi. Pada penelitian ini, limit nilai *coherence* yang peneliti gunakan berjumlah 11. Berikut merupakan grafik nilai *coherence* yang didapatkan untuk pemodelan topik tahun 2020.



Gambar 5.10 Grafik Nilai *Coherence* Tahun 2020

Berdasarkan grafik, diketahui bahwa topik dengan *coherence* tertinggi merupakan *num topics* ke-6. **Gambar 5.10** hanya menampilkan grafik saja, sehingga masih belum diketahui dengan pasti berapa nilai *coherence* yang didapatkan pada jumlah topik tersebut. Sehingga, nilai *coherence* untuk masing-masing jumlah topik di atas dapat dilihat pada **Tabel 5.16**.

Tabel 5.16 Hasil Nilai *Coherence* Tahun 2020

Num Topics	Topic Coherence	Num Topics	Topic Coherence
1	0.25538	6	0.36465
2	0.24852	7	0.34295
3	0.30214	8	0.3432
4	0.30737	9	0.34794
5	0.32976	10	0.35574

Hasil *topic coherence* pada **Tabel 5.16** menunjukkan bahwa *topic coherence* tertinggi adalah *num topics* ke-6 dengan nilai *coherence* sebesar 0.36465. Berdasarkan nilai *coherence* tertinggi, maka jumlah topik tersebutlah yang akan

peneliti gunakan sebagai acuan untuk melakukan analisis *Topic Modeling* Kasus Tindak Pidana Pencurian di Pengadilan Negeri Palembang tahun 2020.

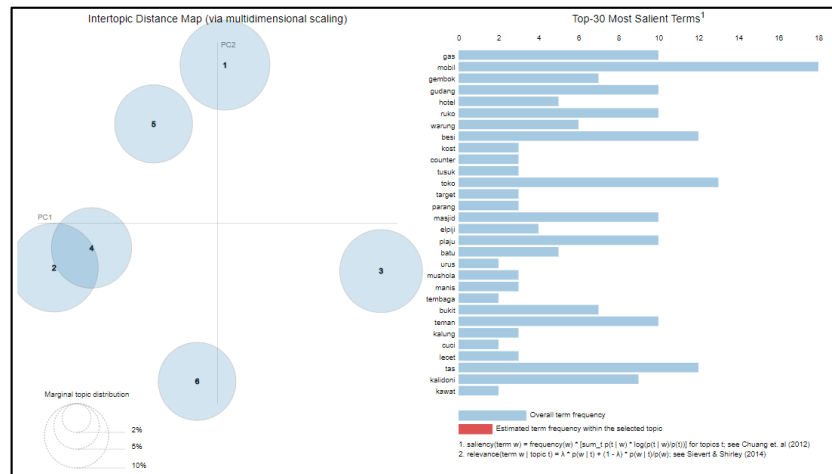
Kemudian, letak dari tiap topik pada titik kordinat dapat dilihat berdasarkan nilai *principal component* (PC) yang terbentuk dari sekian banyaknya frekuensi dokumen yang kemudian direduksi dimensinya dengan membentuk variabel-variabel baru (dalam penelitian ini berupa PC1 dan PC2) serta hanya akan direpresentasikan dalam bentuk diagram kartesius. Masing-masing topik bisa berada di kuadran I, kuadran II, kuadran III, atau kuadran IV tergantung dengan nilai PC yang didapatkan. Kuadran pada visualisasi memiliki fungsi untuk memetakan topik, sehingga akan terlihat sejauh mana topik-topik tersebut saling berdekatan atau terpisah, serta untuk mengetahui pola hubungan dari topik-topik tertentu. Topik yang berdekatan dalam kuadran kemungkinan besar memiliki kemiripan dalam hal kata yang muncul atau topik tersebut berkaitan satu sama lainnya.

Tabel 5.17 Hasil Nilai PC Tahun 2020

Topic	PC1	PC2
1	0.011792	0.045907
2	-0.037062	-0.011810
3	0.056401	-0.012805
4	-0.026345	-0.006198
5	-0.008611	0.029137
6	0.003825	-0.044230

Tabel 5.17 menunjukkan nilai *principal component* (PC) untuk masing-masing topik yang akan digunakan dalam visualisasi *pyLDAvis*. PC1 menunjukkan titik koordinat pada sumbu X, sedangkan nilai PC2 digunakan sebagai titik koordinat pada sumbu Y. Pada topik 1, didapatkan PC1 bernilai positif dan PC2 bernilai positif, sehingga topik 1 akan terletak di kuadran 1. Pada topik 2, didapatkan PC1 bernilai negatif dan PC2 bernilai negatif, sehingga topik 2 akan terletak di kuadran 3. Pada topik 3, didapatkan PC1 bernilai positif dan PC2 bernilai negatif, sehingga topik 3 akan terletak di kuadran 4. Pada topik 4, didapatkan PC1 bernilai negatif dan PC2 bernilai negatif, sehingga topik 4 akan terletak di kuadran 3. Pada topik 5, didapatkan PC1 bernilai negatif dan PC2 bernilai positif, sehingga

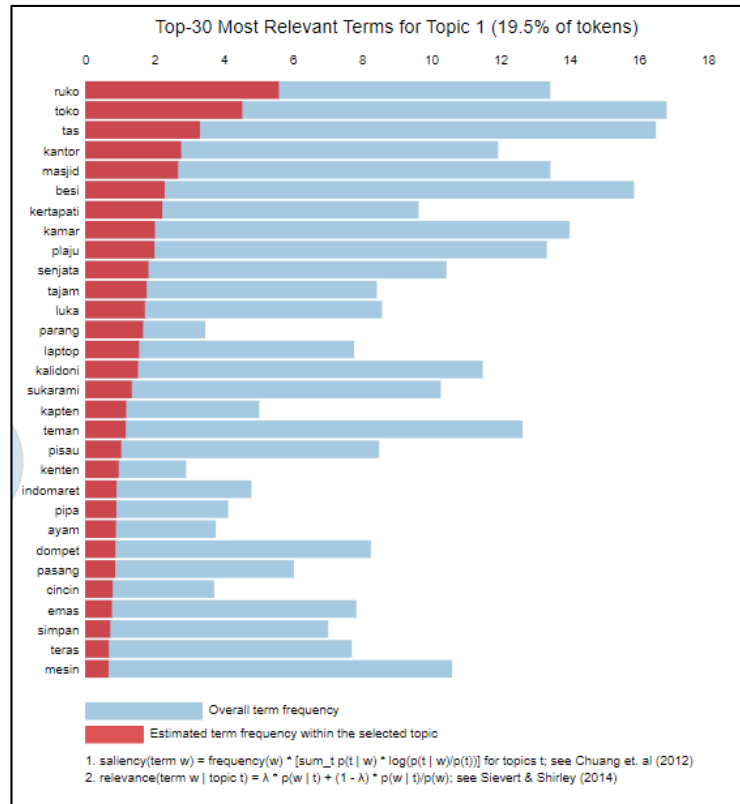
topik 5 akan terletak di kuadran 2. Pada topik 6, didapatkan PC1 bernilai positif dan PC2 bernilai negatif, sehingga topik 6 akan terletak di kuadran 4. Dari **Gambar 5.12**, terlihat bahwa letak dari masing-masing topik sudah sesuai dengan pemosisian berdasarkan nilai PC1 dan PC2 yang telah dijelaskan sebelumnya. Hasil visualisasi dari *pyLDAvis* terlihat seperti berikut.



Gambar 5.11 Visualisasi Keenam Topik Tahun 2020

Berdasarkan **Gambar 5.11** pada bagian kiri terlihat bahwa letak topik 2, topik 4, dan topik 6 berada di kuadran yang sama, serta topik 2 dan topik 4 beririsan, sehingga dapat dikatakan bahwa kedua topik tersebut memiliki hubungan yang kuat. Jika jarak antar topik saling berdekatan maka di dalam kedua topik akan muncul kata-kata yang relatif sama. Sedangkan, letak antara topik 2 dan 4, dengan topik lainnya menyebar pada kuadran yang berbeda-beda, sehingga topik 2 dan 4 dengan topik lainnya memiliki hubungan antar topik yang lemah. Visualisasi di atas juga menunjukkan *bar chart top 30 global term* paling selaras dengan suatu topik tertentu. *Bar chart* berwarna biru menunjukkan keseluruhan *term frequency* dalam dokumen, sementara itu, *bar chart* berwarna merah menunjukkan *term frequency* dalam suatu topik tertentu.

Analisis pemodelan topik bertujuan untuk melihat topik-topik yang terbentuk dari sekumpulan data teks, dengan melihat *output* atau visualisasi yang dihasilkan, lalu menyimpulkan menjadi suatu topik tertentu. Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-1 yang didapatkan:



Gambar 5.12 Visualisasi Topik 1 Tahun 2020

Pada **Gambar 5.12**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 1, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 1 adalah sebesar 19.5%, yang artinya pentingnya topik terhadap dokumen sebesar 19.5%.

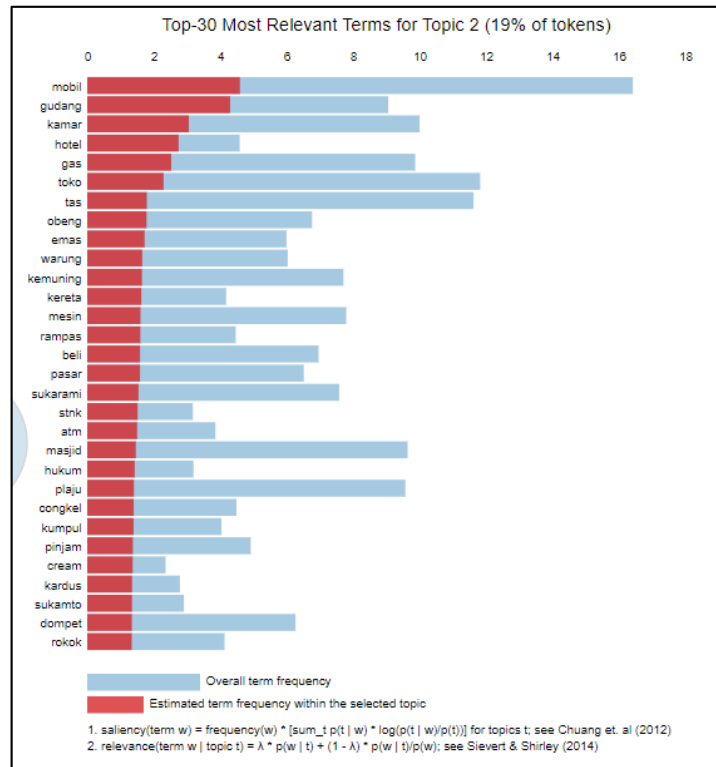
Tabel 5.18 Model Topik 1 Tahun 2020

Model Topik 1
$0.014 * \text{"ruko"} + 0.012 * \text{"toko"} + 0.010 * \text{"tas"} + 0.008 * \text{"kantor"} + 0.008 * \text{"masjid"} + 0.008 * \text{"besi"} + 0.007 * \text{"kertapati"} + 0.007 * \text{"kamar"} + 0.007 * \text{"plaju"} + 0.007 * \text{"senjata"}$

Tabel 5.18 menampilkan model pada topik 1, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 1 dari analisis LDA tahun 2020 membahas mengenai target pencurian. Hal ini didasari kata ‘ruko’, ‘toko’, dan ‘kantor’ sehingga diketahui bahwa target pencurian bukanlah seorang individu. Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, pada kasus ini ketiga terdakwa sepakat melakukan pencurian di suatu **Kantor** Lurah dengan cara merusak pintu, terdakwa berhasil masuk dan mengambil barang-

barang yang ada di dalam **Kantor** Lurah tersebut berupa komputer, printer, televisi, dan lain sebagainya.

Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-2 yang didapatkan:



Gambar 5.13 Visualisasi Topik 2 Tahun 2020

Pada **Gambar 5.13**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 2, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 2 adalah sebesar 19%, yang artinya pentingnya topik terhadap dokumen sebesar 19%.

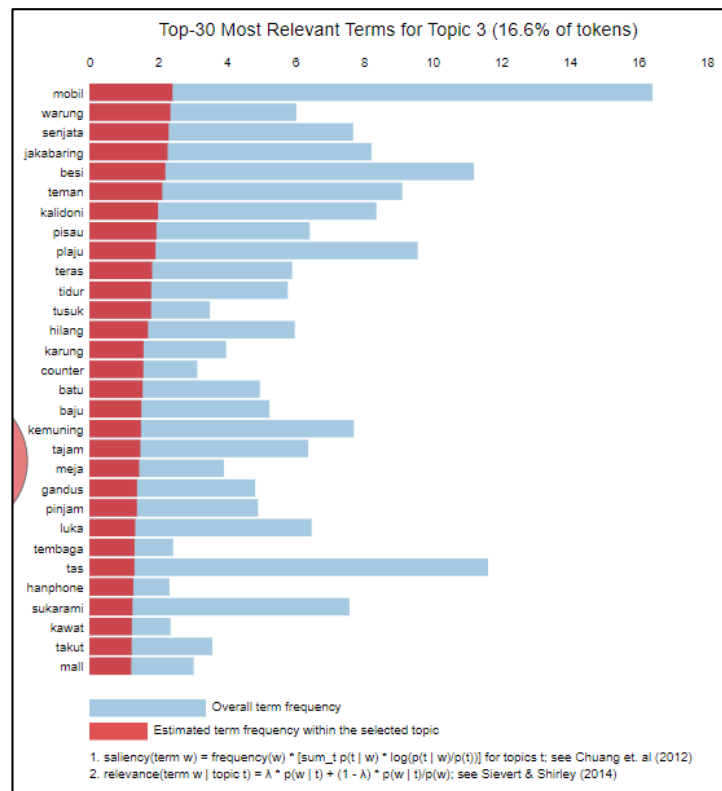
Tabel 5.19 Model Topik 2 Tahun 2020

Model Topik 2
$0.015 * \text{"mobil"} + 0.014 * \text{"gudang"} + 0.010 * \text{"kamar"} + 0.009 * \text{"hotel"} + 0.008 * \text{"gas"} + 0.007 * \text{"toko"} + 0.006 * \text{"tas"} + 0.006 * \text{"obeng"} + 0.006 * \text{"emas"} + 0.005 * \text{"warung"}$

Tabel 5.19 menampilkan model pada topik 2, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 2 dari analisis LDA tahun 2020 membahas mengenai target pencurian. Hal ini didasari kata ‘gudang’, ‘hotel’ dan ‘toko’, sehingga diketahui bahwa target pencurian bukanlah seorang individu.

Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, kasus pencurian ini bertempat di **Hotel X**, dimana para tersangka mengambil barang yang ada di gedung tempat alat-alat *catering* disimpan, benda yang diambil berupa 13.5 lusin sendok steinles, 3 *cavingdish*, dan lain sebagainya.

Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-3 yang didapatkan:



Gambar 5.14 Visualisasi Topik 3 Tahun 2020

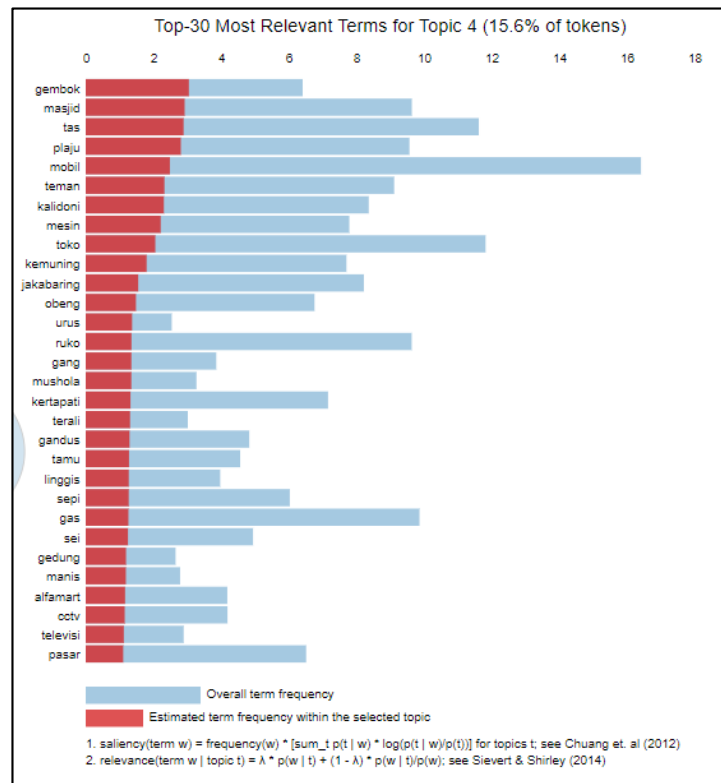
Pada **Gambar 5.14**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 3, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 3 adalah sebesar 16.6%, yang artinya pentingnya topik terhadap dokumen sebesar 16.6%.

Tabel 5.20 Model Topik 3 Tahun 2020

Model Topik 3			
0.009*	"mobil"	+	0.009*
0.009*	"warung"	+	0.009*
0.009*	"senjata"	+	0.009*
0.008*	"jakabaring"	+	0.008*
0.008*	"besi"	+	0.008*
0.008*	"teman"	+	0.008*
0.007*	"kalidoni"	+	0.007*
0.007*	"pisau"	+	0.007*
0.007*	"plaju"	+	0.007*
0.007*	"teras"	+	0.007*

Tabel 5.20 menampilkan model pada topik 3, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 3 dari analisis LDA tahun 2020 membahas mengenai jenis pencurian. Hal ini didasari kata ‘senjata’ dan ‘pisau’ yang mengindikasikan pencurian dengan kekerasan. Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, pada kasus ini terdakwa memanggil korban yang lewat di depannya kemudian terdakwa mengambil *handphone* milik korban, lalu korban mengambil dan merebut kembali barang miliknya tersebut, ketika korban hendak melarikan diri terdakwa langsung mengambil senjata tajam jenis **pisau** miliknya dan **pisau** tersebut ditusukan ke bagian punggung kiri serta dada kiri korban.

Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-4 yang didapatkan:



Gambar 5.15 Visualisasi Topik 4 Tahun 2020

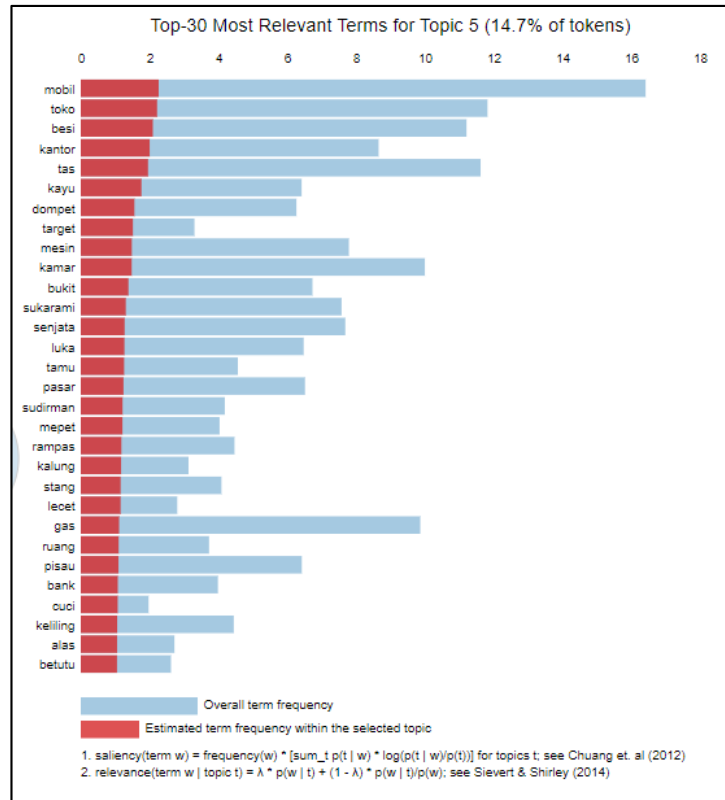
Pada **Gambar 5.15**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 4, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 4 adalah sebesar 15.6%, yang artinya pentingnya topik terhadap dokumen sebesar 15.6%.

Tabel 5.21 Model Topik 4 Tahun 2020

Model Topik 4
<code>\0.012*"gembok" + 0.011*"masjid" + 0.011*"tas" + 0.01*"plaju" + 0.010*"mobil" + 0.009*"teman" + 0.009*"kalidoni" + 0.009*"mesin" + 0.008*"toko" + 0.005*"kemuning"</code>

Tabel 5.21 menampilkan model pada topik 4, berdasarkan hasil model yang didapatkan, terlihat ada 3 nama daerah yaitu ‘plaju’, ‘kalidoni’, dan ‘kemuning’ dari ketiga daerah tersebut mengindikasikan 3 daerah yang banyak terjadi kasus pencurian. Sehingga, dapat peneliti simpulkan bahwa untuk topik 4 dari analisis LDA tahun 2020 membahas mengenai daerah tempat terjadinya pencurian. Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, kasus pertama ini terjadi di Kecamatan **Kemuning** dimana terdakwa mengambil 1 unit sepeda motor yang sedang terparkir di depan sebuah toko. Kasus kedua terjadi di Kecamatan **Plaju**, terdakwa memasuki rumah korban yang pintu pagarnya terbuka dan masuk ke rumah tersebut melalui pintu lantai 2 yang tidak terkunci, kemudian terdakwa mengambil 1 buah tas merk Vinlee yang di dalamnya terdapat 1 unit *handphone*, dompet, 2 buah gelang 18 karat, dan lain sebagainya.

Berikut ini merupakan visualisasi dengan menggunakan *pyLDavis* dari topik ke-5 yang didapatkan:



Gambar 5.16 Visualisasi Topik 5 Tahun 2020

Pada **Gambar 5.16**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 5, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 5 adalah sebesar 14.7%, yang artinya pentingnya topik terhadap dokumen sebesar 14.7%.

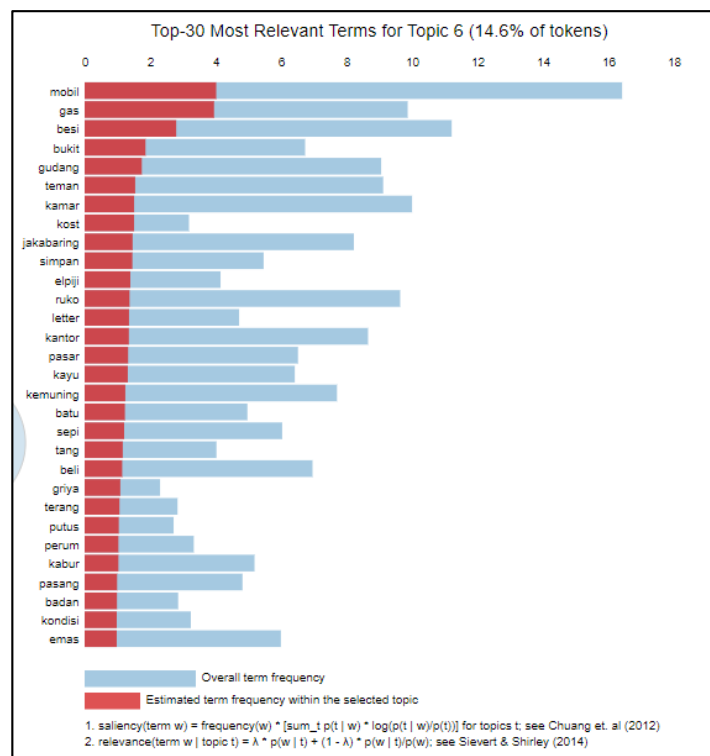
Tabel 5.22 Model Topik 5 Tahun 2020

Model Topik 5
$0.009 * "mobil" + 0.009 * "toko" + 0.009 * "besi" + 0.008 * "kantor" + 0.008 * "tas" + 0.007 * "kayu" + 0.006 * "dompet" + 0.006 * "target" + 0.006 * "mesin" + 0.006 * "kamar"$

Tabel 5.22 menampilkan model pada topik 5, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 5 dari analisis LDA tahun 2020 membahas mengenai alat pendukung pencurian. Hal ini didasari kata ‘besi’, dan ‘kayu’ yang mengindikasikan bahwa kedua alat tersebut digunakan pelaku untuk mempermudah melaksanakan aksinya. Kemudian, peneliti juga mencari kata tersebut pada data surat dakwaan dan peneliti menemukan bahwa benar kata “besi” dan “kayu” merupakan alat yang digunakan para terdakwa untuk membantu

menjalankan aksi pencurian. Contohnya, pada suatu kasus terdakwa menggunakan 1 buah **kayu** untuk mencongkel pagar teralis toko. Pada kasus lainnya, seorang terdakwa menggunakan sebuah gunting **besi** untuk memotong engsel kunci rumah korban, lalu mengambil 1 unit laptop ASUS yang terletak di ruang tengah rumah tersebut.

Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-6 yang didapatkan:



Gambar 5.17 Visualisasi Topik 6 Tahun 2020

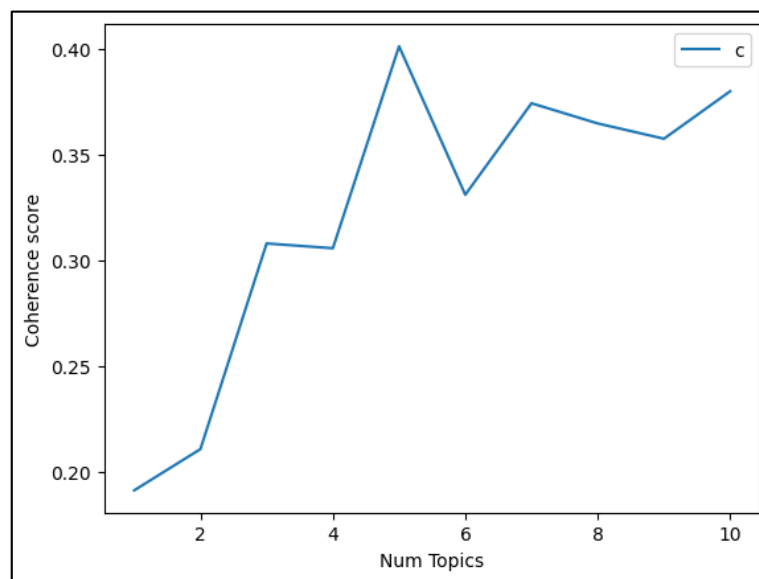
Pada **Gambar 5.17**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 6, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 6 adalah sebesar 14.6%, yang artinya pentingnya topik terhadap dokumen sebesar 14.6%.

Tabel 5.23 Model Topik 6 Tahun 2020

Model Topik 6
<code>\0.009*"toko" + 0.007*"kantor" + 0.007*"bukit" + 0.007*"yamaha" + 0.006*"samsung" + 0.006*"alfamart" + 0.006*"plaju" + 0.006*"ruko" + 0.006*"hotel" + 0.006*"kalidoni"</code>

5.4.3 Tahun 2021

Untuk menentukan jumlah topik dalam analisis *Topic Modeling* dapat dilakukan dengan melihat visualisasi pada grafik *coherence score*. *Coherence score* adalah nilai yang digunakan untuk mengukur seberapa baik pemodelan topik, semakin tinggi nilai *coherence* maka akan semakin baik pula model yang akan didapatkan. Sehingga, yang akan digunakan dalam analisis *Topic Modeling* adalah jumlah topik dengan nilai *coherence* tertinggi. Pada penelitian ini, limit nilai *coherence* yang peneliti gunakan berjumlah 11. Berikut merupakan grafik nilai *coherence* yang didapatkan untuk pemodelan topik tahun 2021.



Gambar 5.19 Grafik Nilai *Coherence* Tahun 2021

Berdasarkan grafik, diketahui bahwa topik dengan *coherence* tertinggi merupakan *num topics* ke-5. **Gambar 5.19** hanya menampilkan grafik saja, sehingga masih belum diketahui dengan pasti berapa nilai *coherence* yang didapatkan pada jumlah topik tersebut. Sehingga, nilai *coherence* untuk masing-masing jumlah topik di atas dapat dilihat pada **Tabel 5.24**.

Tabel 5.24 Hasil Nilai *Coherence* Tahun 2021

Num Topics	Topic Coherence	Num Topics	Topic Coherence
1	0.19135	6	0.33105
2	0.21084	7	0.37427
3	0.30803	8	0.36474
4	0.30576	9	0.35751

5	0.40121	10	0.37995
---	---------	----	---------

Hasil *topic coherence* pada **Tabel 5.24** menunjukkan bahwa *topic coherence* tertinggi adalah *num topics* ke-5 dengan nilai *coherence* sebesar 0.40121. Berdasarkan nilai *coherence* tertinggi, maka jumlah topik tersebutlah yang akan peneliti gunakan sebagai acuan untuk melakukan analisis *Topic Modeling* Kasus Tindak Pidana Pencurian di Pengadilan Negeri Palembang tahun 2021.

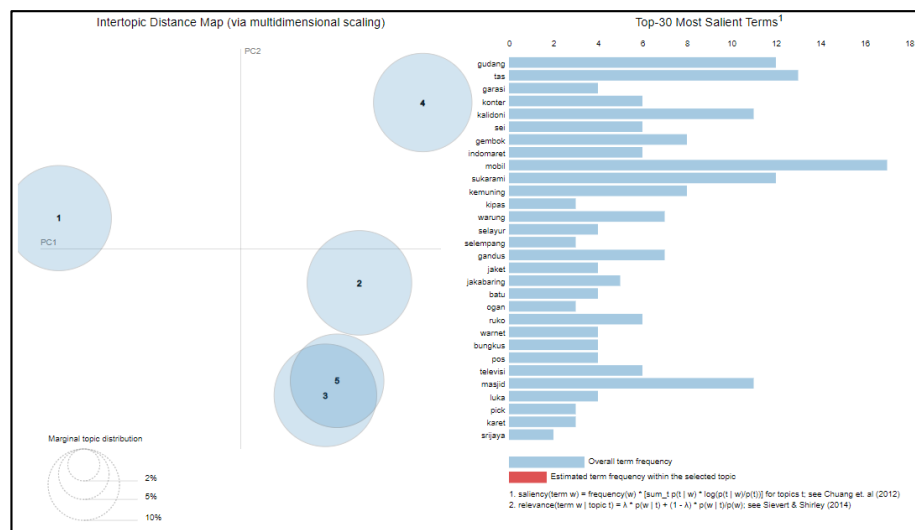
Kemudian, letak dari tiap topik pada titik kordinat dapat dilihat berdasarkan nilai *principal component* (PC) yang terbentuk dari sekian banyaknya frekuensi dokumen yang kemudian direduksi dimensinya dengan membentuk variabel-variabel baru (dalam penelitian ini berupa PC1 dan PC2) serta hanya akan direpresentasikan dalam bentuk diagram kartesius. Masing-masing topik bisa berada di kuadran I, kuadran II, kuadran III, atau kuadran IV tergantung dengan nilai PC yang didapatkan. Kuadran pada visualisasi memiliki fungsi untuk memetakan topik, sehingga akan terlihat sejauh mana topik-topik tersebut saling berdekatan atau terpisah, serta untuk mengetahui pola hubungan dari topik-topik tertentu. Topik yang berdekatan dalam kuadran kemungkinan besar memiliki kemiripan dalam hal kata yang muncul atau topik tersebut berkaitan satu sama lainnya.

Tabel 5.25 Hasil Nilai PC Tahun 2021

Topic	PC1	PC2
1	-0.066318	0.015862
2	0.016101	-0.001897
3	0.006773	-0.032756
4	0.033460	0.047568
5	0.009984	-0.028776

Tabel 5.25 menunjukan nilai *principal component* (PC) untuk masing-masing topik yang akan digunakan dalam visualisasi *pyLDavis*. PC1 menunjukan titik koordinat pada sumbu X, sedangkan nilai PC2 digunakan sebagai titik koordinat pada sumbu Y. Pada topik 1, didapatkan PC1 bernilai negatif dan PC2 bernilai positif, sehingga topik 1 akan terletak di kuadran 2. Pada topik 2, didapatkan PC1 bernilai positif dan PC2 bernilai negatif, sehingga topik 2 akan terletak di kuadran 4. Pada topik 3, didapatkan PC1 bernilai positif dan PC2 bernilai

negatif, sehingga topik 3 akan terletak di kuadran 4. Pada topik 4, didapatkan PC1 bernilai positif dan PC2 bernilai positif, sehingga topik 4 akan terletak di kuadran 1. Pada topik 5, didapatkan nilai PC1 bernilai positif dan PC2 bernilai negatif, sehingga topik 5 akan terletak di kuadran 4. Dari **Gambar 5.20**, terlihat bahwa letak dari masing-masing topik sudah sesuai dengan pemosisian berdasarkan nilai PC1 dan PC2 yang telah dijelaskan sebelumnya. Hasil visualisasi dari *pyLDAvis* terlihat seperti berikut.

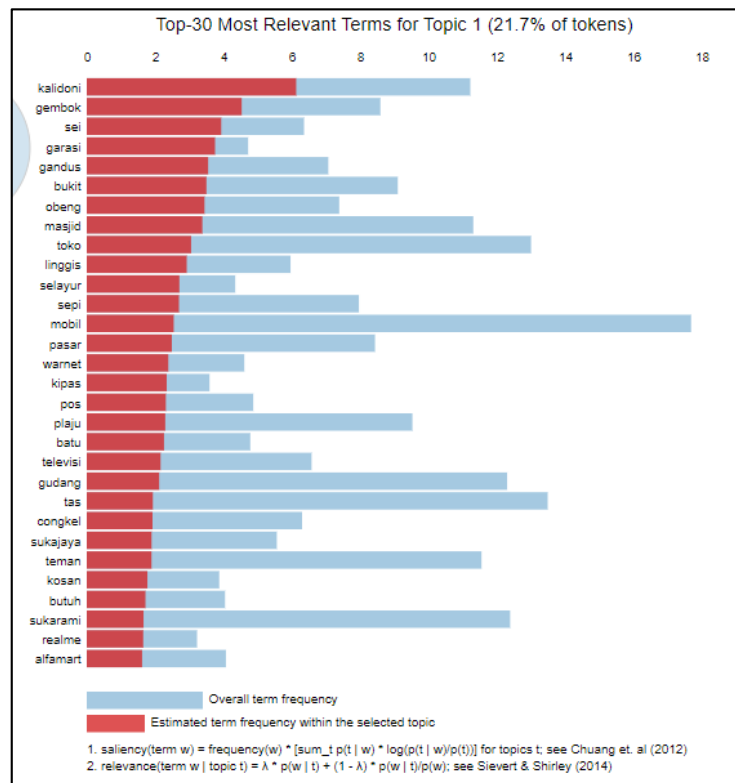


Gambar 5.20 Visualisasi Kelima Topik Tahun 2021

Berdasarkan **Gambar 5.20** pada bagian kiri terlihat bahwa letak topik 2, topik 3, dan topik 5 berada di kuadran yang sama serta topik 3 dan topik 5 yang beririsan, sehingga dapat dikatakan bahwa ketiga topik tersebut memiliki hubungan yang kuat. Jika jarak antar topik saling berdekatan maka di dalam ketiga topik akan muncul kata-kata yang relatif sama. Sedangkan, letak antara topik 1, dan topik 4 dengan topik 2, topik 3, dan topik 5 menyebar pada kuadran yang berbeda-beda, topik-topik tersebut memiliki hubungan antar topik yang lemah. Visualisasi di atas juga menunjukkan *bar chart top 30 global term* paling selaras dengan suatu topik tertentu. *Bar chart* berwarna biru menunjukkan keseluruhan *term frequency* dalam dokumen, sementara itu, *bar chart* berwarna merah menunjukkan *term frequency* dalam suatu topik tertentu.

Analisis pemodelan topik bertujuan untuk melihat topik-topik yang terbentuk dari sekumpulan data teks, dengan melihat *output* atau visualisasi yang dihasilkan,

lalu menyimpulkan menjadi suatu topik tertentu. Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-1 yang didapatkan:



Gambar 5.21 Visualisasi Topik 1 Tahun 2021

Pada **Gambar 5.22**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 1, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 1 adalah sebesar 21.7%, yang artinya pentingnya topik terhadap dokumen sebesar 21.7%.

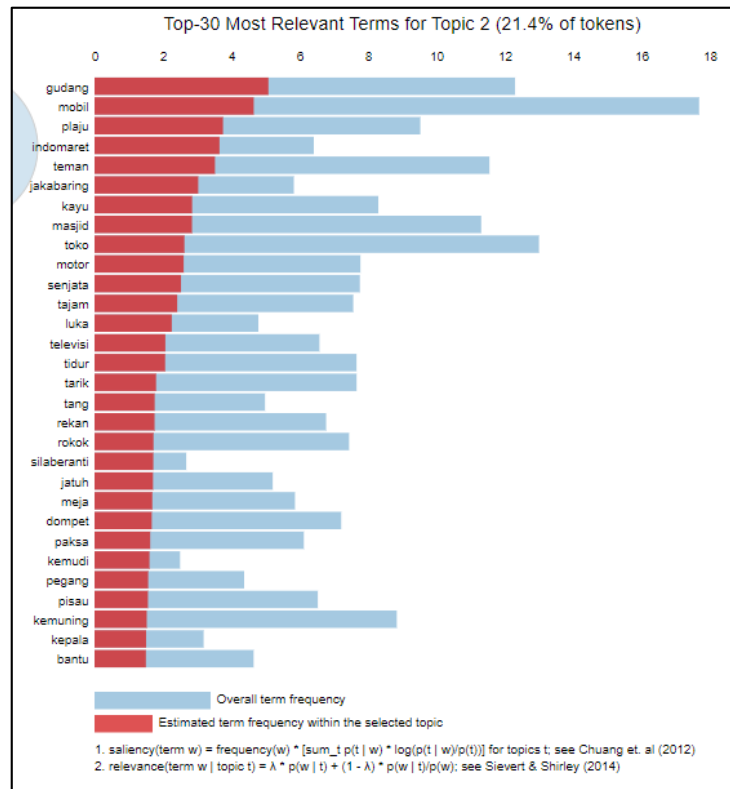
Tabel 5.26 Model Topik 1 Tahun 2021

Model Topik 1
$0.020 * "kalidoni" + 0.015 * "gembok" + 0.013 * "sei" + 0.012 * "garasi"$ $+ 0.011 * "gandum" + 0.011 * "bukit" + 0.011 * "obeng" + 0.011 * "masjid"$ $+ 0.010 * "toko" + 0.009 * "linggis"$

Tabel 5.26 menampilkan model pada topik 1, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 1 dari analisis LDA tahun 2021 membahas mengenai alat pendukung pencurian. Hal ini didasari kata ‘gembok’, ‘obeng’ dan ‘linggis’ yang mengindikasikan alat yang mendukung pelaku untuk melancarkan aksinya. Berikut contoh kasus pada Surat Dakwaan yang

menyebutkan salah satu kata di atas, diketahui pada suatu kasus, linggis digunakan untuk merusak pintu, selain itu pada kasus lain, linggis juga digunakan untuk mencongkel daun jendela agar dapat masuk ke rumah korban.

Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-2 yang didapatkan:



Gambar 5.22 Visualisasi Topik 2 Tahun 2021

Pada **Gambar 5.22**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 2, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 2 adalah sebesar 21.4%, yang artinya pentingnya topik terhadap dokumen sebesar 21.4%.

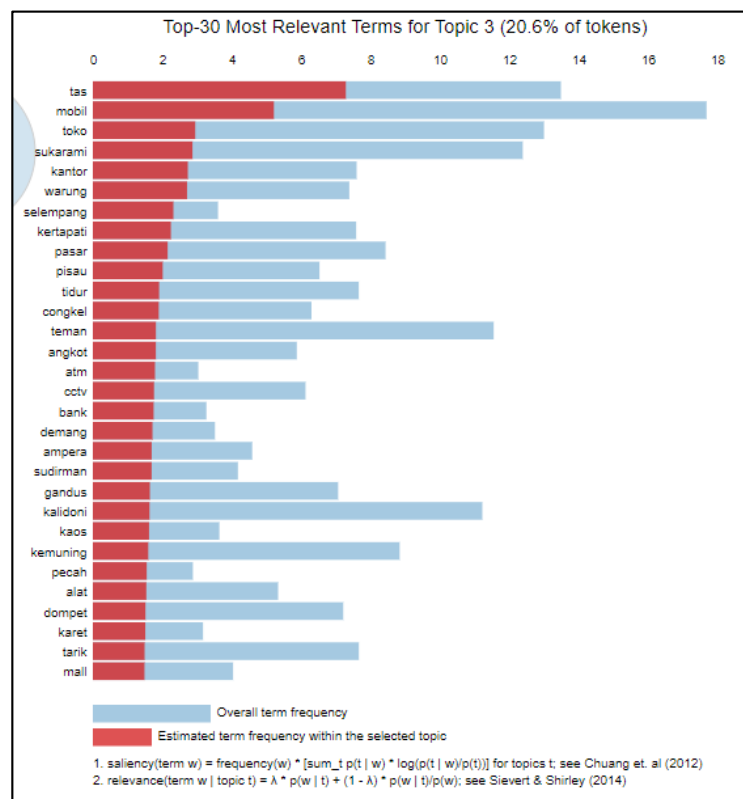
Tabel 5.27 Model Topik 2 Tahun 2021

Model Topik 2			
0.017*	"gudang"	+	0.015*
	"mobil"	+	0.012*
	"plaju"	+	
0.012*	"indomaret"	+	0.012*
	"teman"	+	0.010*
	"jakabaring"	+	
0.009*	"kayu"	+	0.009*
	"masjid"	+	0.009*
	"toko"	+	0.009*
	"motor"		

Tabel 5.27 menampilkan model pada topik 2, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 2 dari analisis LDA tahun

2020 membahas mengenai jenis pencurian. Hal ini didasari kata ‘mobil’, dan ‘motor’, yang mengindikasikan curanmor-R4 dan curanmor-R2. Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, pada kasus ini terdakwa memasuki rumah korban melalui jendela rumahnya yang terbuka tanpa adanya teralis pada jendela, lalu membuka pintu depan rumah korban dengan kunci yang melekat di pintu, kemudian terdakwa mengambil 1 unit sepeda **motor** merk Supra X milik korban.

Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-3 yang didapatkan:



Gambar 5.23 Visualisasi Topik 3 Tahun 2021

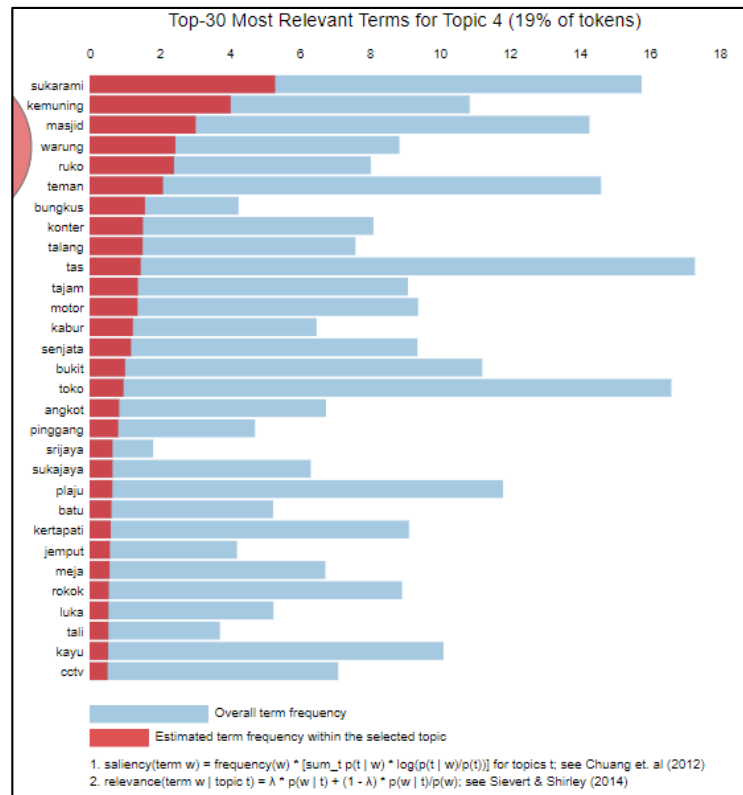
Pada **Gambar 5.24**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 3, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 3 adalah sebesar 20.6%, yang artinya pentingnya topik terhadap dokumen sebesar 20.6%.

Tabel 5.28 Model Topik 3 Tahun 2021

Model Topik 3
$0.025 * \text{"tas"} + 0.018 * \text{"mobil"} + 0.010 * \text{"toko"} + 0.010 * \text{"sukarami"} + 0.009 * \text{"kantor"} + 0.009 * \text{"warung"} + 0.008 * \text{"selempang"} + 0.008 * \text{"kertapati"} + 0.007 * \text{"pasar"} + 0.008 * \text{"pisau"}$

Tabel 5.28 menampilkan model pada topik 3, berdasarkan hasil model yang didapatkan berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 3 dari analisis LDA tahun 2021 membahas mengenai jenis barang curian. Hal ini didasari kata ‘tas’, dan ‘selempang’ yang mengindikasikan barang yang dicuri, dimana kata tas dan selempang juga berkaitan satu sama lainnya. Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, pada kasus ini terdakwa yang melihat **tas selempang** yang digunakan korban, lalu menarik paksa tas milik korban yang berisikan 1 unit Iphone 7+ dan dompet berwarna hitam.

Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-4 yang didapatkan:



Gambar 5.24 Visualisasi Topik 4 Tahun 2021

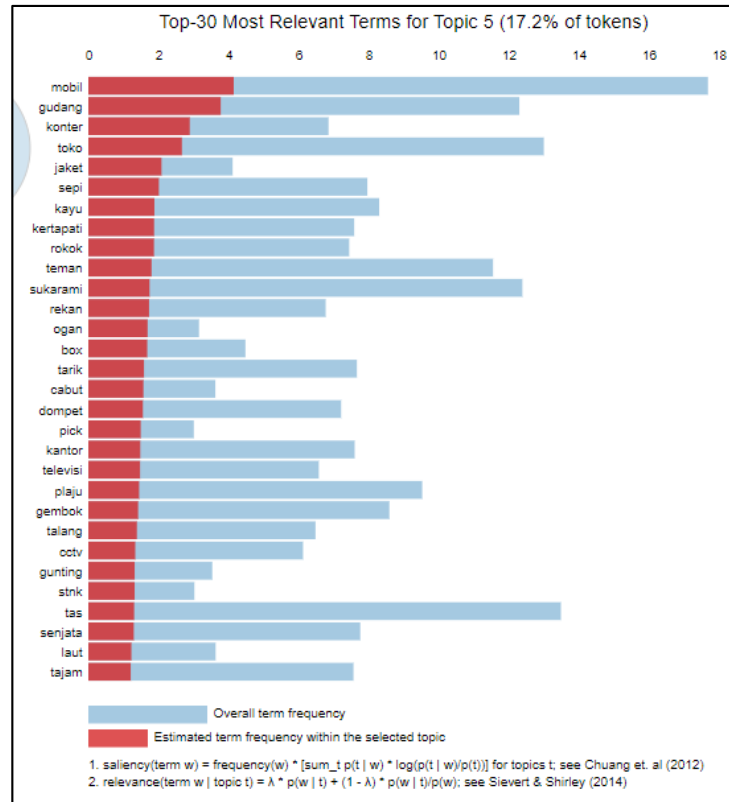
Pada **Gambar 5.24**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 4, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 4 adalah sebesar 19%, yang artinya pentingnya topik terhadap dokumen sebesar 19%.

Tabel 5.29 Model Topik 4 Tahun 2021

Model Topik 4	
<code>\0.018*"sukarami" + 0.014*"kemuning" + 0.012*"masjid" +</code> <code>0.010*"warung" + 0.010*"ruko" + 0.009*"teman" + 0.008*"bungkus" +</code> <code>0.008*"konter" + 0.008*"talang" + 0.008*"tas"</code>	

Tabel 5.29 menampilkan model pada topik 4, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 4 dari analisis LDA tahun 2021 membahas mengenai tempat terjadinya pencurian. Hal ini didasari kata “masjid” dan “warung”. Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, pada kasus ini terjadi di sebuah **masjid**, dimana terdakwa merusak pintu **masjid** tersebut dengan menggunakan obeng, lalu mengambil kipas angin gantung, facum cleaner milik **masjid**.

Berikut ini merupakan visualisasi dengan menggunakan *pyLDavis* dari topik ke-5 yang didapatkan:



Gambar 5.25 Visualisasi Topik 5 Tahun 2021

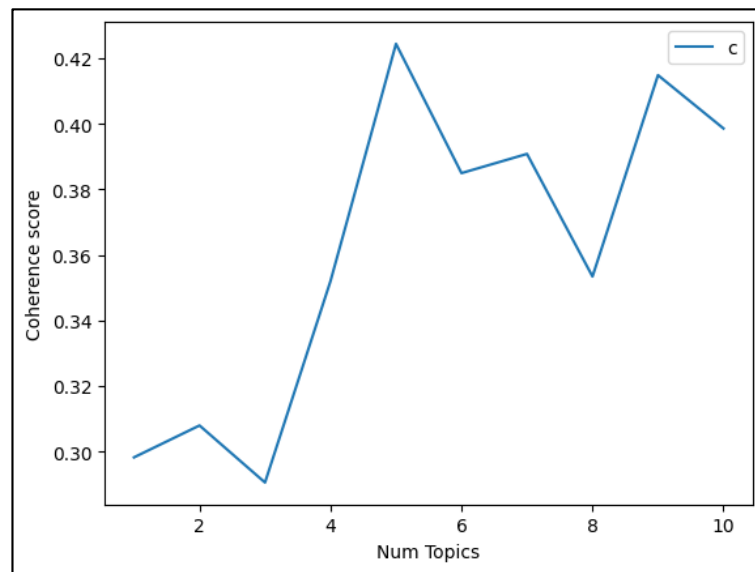
Pada **Gambar 5.25**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 5, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 5 adalah sebesar 17.2%, yang artinya pentingnya topik terhadap dokumen sebesar 17.2%.

Tabel 5.30 Model Topik 5 Tahun 2021

Model Topik 5
$0.017 * "mobil" + 0.015 * "gudang" + 0.012 * "konter" + 0.011 * "toko" + 0.009 * "jaket" + 0.008 * "sepi" + 0.008 * "kayu" + 0.008 * "kertapati" + 0.008 * "rokok" + 0.007 * "teman"$

Tabel 5.30 menampilkan model pada topik 5, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 5 dari analisis LDA tahun 2021 membahas mengenai target pencurian. Hal ini didasari kata ‘gudang’, ‘konter’, dan ‘toko’ sehingga diketahui bahwa target pencurian bukanlah seorang individu. Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, pada kasus ini terdakwa menargetkan sebuah konter *handphone* di kompleks mall X, dimana terdakwa membuka paksa etalase konter tersebut dan

coherence yang peneliti gunakan berjumlah 11. Berikut merupakan grafik nilai *coherence* yang didapatkan untuk pemodelan topik tahun 2022.



Gambar 5.27 Grafik Nilai *Coherence* Tahun 2022

Berdasarkan grafik, diketahui bahwa topik dengan *coherence* tertinggi merupakan *num topics* ke-5. **Gambar 5.27** hanya menampilkan grafik saja, sehingga masih belum diketahui dengan pasti berapa nilai *coherence* yang didapatkan pada jumlah topik tersebut. Sehingga, nilai *coherence* untuk masing-masing jumlah topik di atas dapat dilihat pada **Tabel 5.31**

Tabel 5.31 Hasil Nilai *Coherence* Tahun 2022

Num Topics	Topic Coherence	Num Topics	Topic Coherence
1	0.29828	6	0.38498
2	0.308	7	0.39084
3	0.29061	8	0.35345
4	0.35205	9	0.41486
5	0.42441	10	0.3986

Hasil *topic coherence* pada **Tabel 5.31** menunjukkan bahwa *topic coherence* tertinggi adalah *num topics* ke-5 dengan nilai *coherence* sebesar 0.42441. Berdasarkan nilai *coherence* tertinggi, maka jumlah topik tersebutlah yang akan peneliti gunakan sebagai acuan untuk melakukan analisis *Topic Modeling* Kasus Tindak Pidana Pencurian di Pengadilan Negeri Palembang tahun 2022.

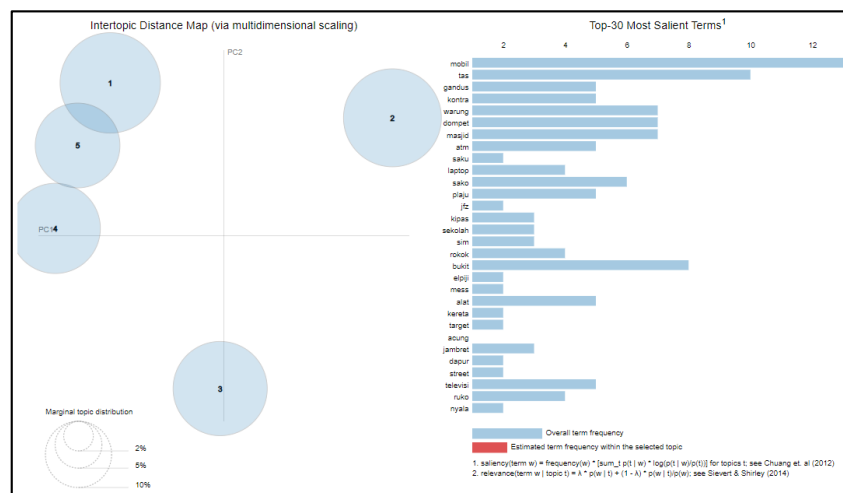
Kemudian, letak dari tiap topik pada titik kordinat dapat dilihat berdasarkan nilai *principal component* (PC) yang terbentuk dari sekian banyaknya frekuensi dokumen yang kemudian direduksi dimensinya dengan membentuk variabel-variabel baru (dalam penelitian ini berupa PC1 dan PC2) serta hanya akan direpresentasikan dalam bentuk diagram kartesius. Masing-masing topik bisa berada di kuadran I, kuadran II, kuadran III, atau kuadran IV tergantung dengan nilai PC yang didapatkan. Kuadran pada visualisasi memiliki fungsi untuk memetakan topik, sehingga akan terlihat sejauh mana topik-topik tersebut saling berdekatan atau terpisah, serta untuk mengetahui pola hubungan dari topik-topik tertentu. Topik yang berdekatan dalam kuadran kemungkinan besar memiliki kemiripan dalam hal kata yang muncul atau topik tersebut berkaitan satu sama lainnya.

Tabel 5.32 Hasil Nilai PC Tahun 2022

Topic	PC1	PC2
1	-0.013796	0.024840
2	0.050124	0.016932
3	0.011103	-0.044347
4	-0.026244	-0.008096
5	-0.021188	0.010671

Tabel 5.32 menunjukkan nilai *principal component* (PC) untuk masing-masing topik yang akan digunakan dalam visualisasi PyLDAvis. PC1 menunjukkan titik koordinat pada sumbu X, sedangkan nilai PC2 digunakan sebagai titik koordinat pada sumbu Y. Pada topik 1, didapatkan PC1 bernilai negatif dan PC2 bernilai positif, sehingga topik 1 akan terletak di kuadran 2. Pada topik 2, didapatkan PC1 bernilai positif dan PC2 bernilai positif, sehingga topik 2 akan terletak di kuadran 1. Pada topik 3, didapatkan PC1 bernilai positif dan PC2 bernilai negatif, sehingga topik 3 akan terletak di kuadran 4. Pada topik 4, didapatkan PC1 bernilai negatif dan PC2 bernilai negatif, sehingga topik 4 akan terletak di kuadran 3. Pada topik 5, didapatkan PC1 bernilai negatif dan PC2 bernilai positif, sehingga topik 5 akan terletak di kuadran 2. Dari **Gambar 5.29**, terlihat bahwa letak dari masing-masing topik sudah sesuai dengan pemosisian berdasarkan nilai PC1 dan

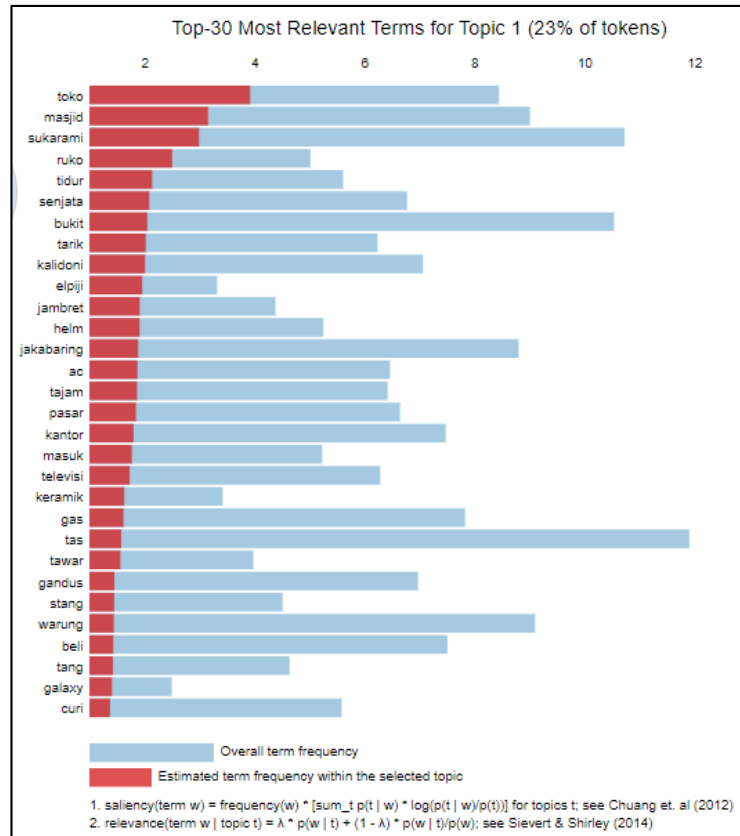
PC2 yang telah dijelaskan sebelumnya. Hasil visualisasi dari *pyLDAvis* terlihat seperti berikut.



Gambar 5.28 Visualisasi Kelima Topik Tahun 2022

Berdasarkan **Gambar 5.28** pada bagian kiri terlihat bahwa letak topik 1 beririsan dengan topik 5, dan topik 5 beririsan dengan topik 4, sehingga dapat dikatakan bahwa masing-masing topik tersebut memiliki hubungan yang cukup kuat. Jika jarak antar topik saling berdekatan maka di dalam ketiga topik akan muncul kata-kata yang relatif sama. Sedangkan, letak antara topik 1, topik 4, dan topik 5 dengan topik 2, serta topik 3, menyebar pada kuadran yang berbeda-beda, sehingga topik-topik tersebut memiliki hubungan antar topik yang lemah. Visualisasi di atas juga menunjukkan *bar chart top 30 global term* paling selaras dengan suatu topik tertentu. *Bar chart* berwarna biru menunjukkan keseluruhan *term frequency* dalam dokumen, sementara itu, *bar chart* berwarna merah menunjukkan *term frequency* dalam suatu topik tertentu.

Analisis pemodelan topik bertujuan untuk melihat topik-topik yang terbentuk dari sekumpulan data text, dengan melihat *output* atau visualisasi yang dihasilkan, lalu menyimpulkan menjadi suatu topik tertentu. Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-1 yang didapatkan:



Gambar 5.29 Visualisasi Topik 1 Tahun 2022

Pada **Gambar 5.29**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 1, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 1 adalah sebesar 21.9%, yang artinya pentingnya topik terhadap dokumen sebesar 21.9%.

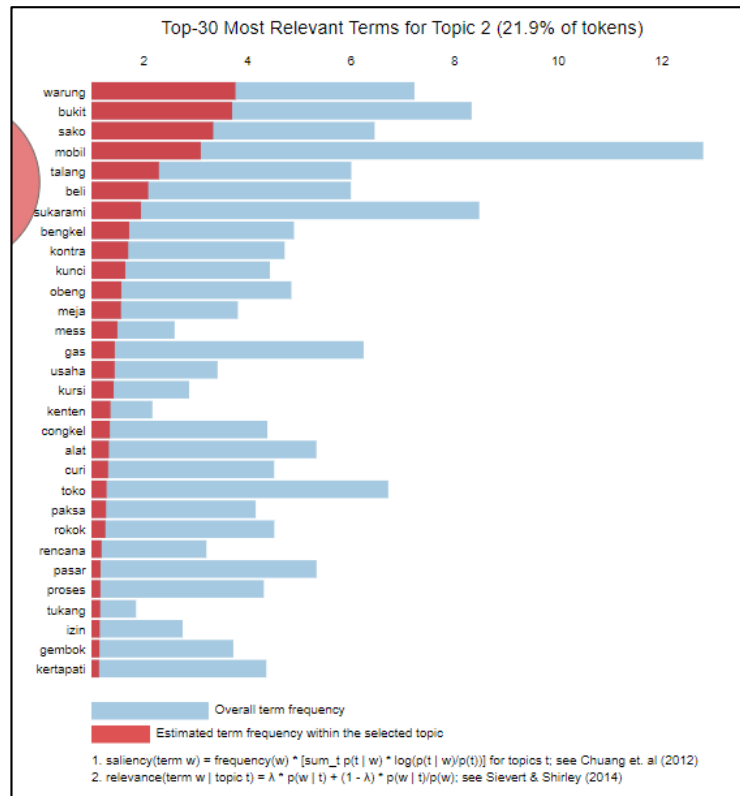
Tabel 5.33 Model Topik 1 Tahun 2022

Model Topik 1
$\begin{aligned} & \backslash 0.012 * \text{"toko"} + 0.010 * \text{"masjid"} + 0.009 * \text{"sukarami"} + 0.008 * \text{"ruko"} \\ & + 0.007 * \text{"tidur"} + 0.007 * \text{"senjata"} + 0.007 * \text{"bukit"} + 0.007 * \text{"tarik"} \\ & + 0.007 * \text{"kalidoni"} + 0.006 * \text{"elpiji"} \end{aligned}$

Tabel 5.33 menampilkan model pada topik 1, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 1 dari analisis LDA tahun 2022 membahas mengenai target pencurian. Hal ini didasari kata ‘toko’, dan ‘ruko’ sehingga diketahui bahwa target pencurian bukanlah seorang individu. Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, pada kasus ini terdakwa menargetkan sebuah **toko X**, dimana terdakwa berhasil

mengambil besi pipa sepanjang setengah meter sebanyak 30 kg yang berada di dalam gudang bagian depan **toko** tersebut, selain itu terdakwa mengambil seng Alma sebanyak 6 kg yang berada di ruang tengah **toko** dengan menggunakan 1 buah linggis yang telah terdakwa siapkan.

Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-2 yang didapatkan:



Gambar 5.30 Visualisasi Topik 2 Tahun 2022

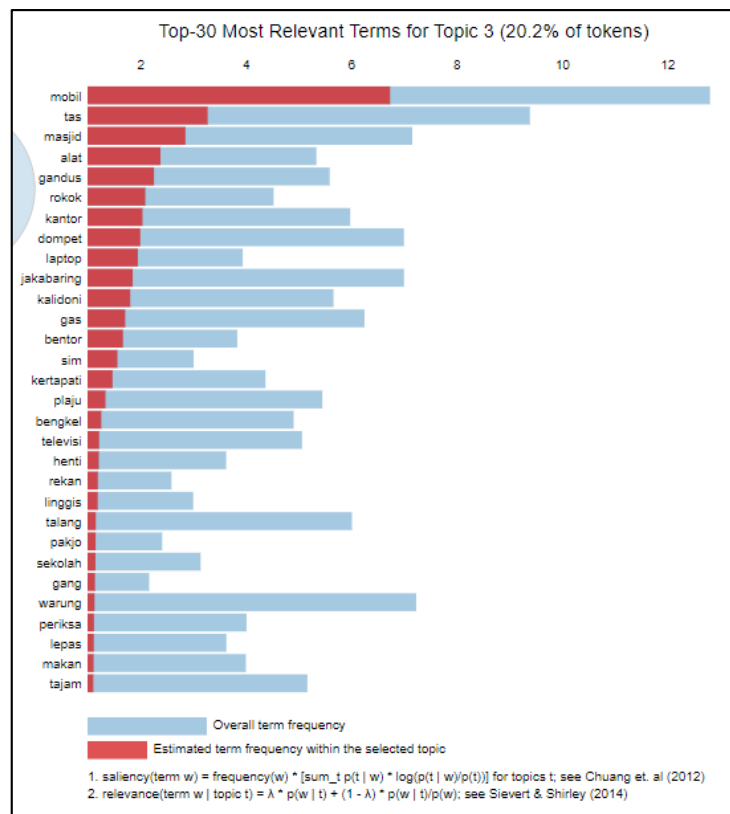
Pada **Gambar 5.30**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 2, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 2 adalah sebesar 21.9%, yang artinya pentingnya topik terhadap dokumen sebesar 21.9%.

Tabel 5.34 Model Topik 2 Tahun 2022

Model Topik 2
$\backslash 0.015 * \text{"warung"} + 0.015 * \text{"bukit"} + 0.013 * \text{"sako"} + 0.012 * \text{"mobil"} + 0.009 * \text{"talang"} + 0.008 * \text{"beli"} + 0.008 * \text{"sukarami"} + 0.008 * \text{"bengkel"} + 0.007 * \text{"kontra"} + 0.006 * \text{"kunci"}$

Tabel 5.34 menampilkan model pada topik 2, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 2 dari analisis LDA tahun 2022 membahas mengenai tempat terjadinya pencurian. Hal ini didasari kata ‘warung’, dan ‘bengkel’. Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, kasus kasus pertama terjadi di sebuah **bengkel**, dimana terdakwa memasuki **bengkel** milik korban yang pintu dan pagarnya tidak terkunci, setelah terdakwa berhasil mengambil 1 set Deksel mobil Toyota Kijang, terdakwa langsung keluar dari **bengkel** milik korban.

Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-3 yang didapatkan:



Gambar 5.31 Visualisasi Topik 3 Tahun 2022

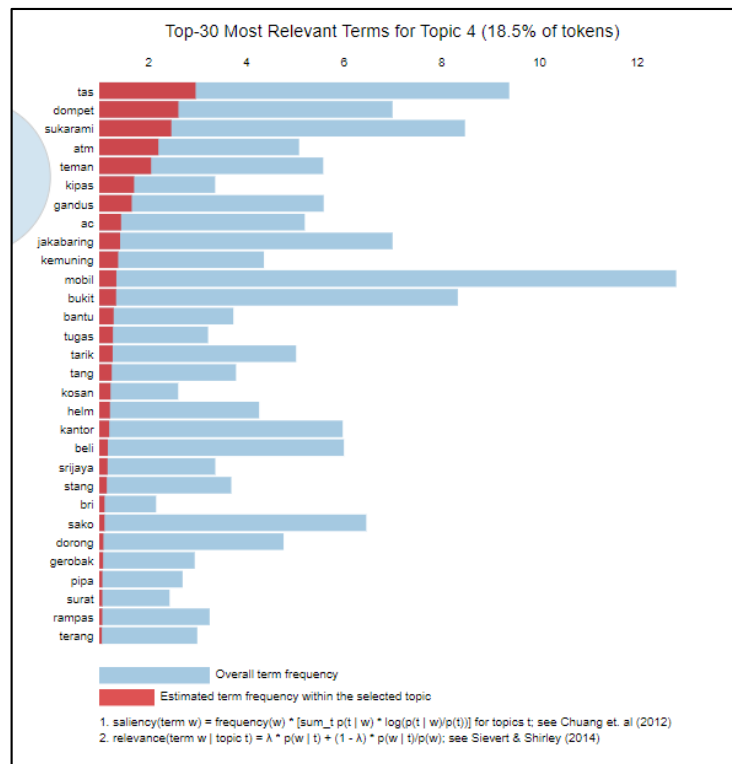
Pada **Gambar 5.31**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 3, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 3 adalah sebesar 20.2%, yang artinya pentingnya topik terhadap dokumen sebesar 20.2%.

Tabel 5.35 Model Topik 3 Tahun 2022

Model Topik 3
$0.029 * \text{"mobil"} + 0.014 * \text{"tas"} + 0.012 * \text{"masjid"} + 0.010 * \text{"alat"} + 0.010 * \text{"gandus"} + 0.009 * \text{"rokok"} + 0.009 * \text{"kantor"} + 0.008 * \text{"dompet"} + 0.008 * \text{"laptop"} + 0.008 * \text{"jakabaring"}$

Tabel 5.35 menampilkan model pada topik 3, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 3 dari analisis LDA tahun 2022 membahas mengenai jenis barang curian. Hal ini didasari kata ‘tas’, ‘rokok’, ‘dompet’, dan ‘laptop’. Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, pada kasus pertama, terdakwa mencuri di suatu warung adapun barang yang dicuri terdiri dari beberapa karung beras, 3 kaleng rokok Surya, 1 unit *handphone* merk samsung warna putih. Kasus kedua terjadi di sebuah rumah, dimana terdakwa mengambil 1 unit **laptop**, 2 buah **tas** kerja yang berisi surat-surat dan 2 buah kunci motor,

Berikut ini merupakan visualisasi dengan menggunakan *pyLDavis* dari topik ke-4 yang didapatkan:



Gambar 5.32 Visualisasi Topik 4 Tahun 2022

Pada **Gambar 5.32**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 4, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai

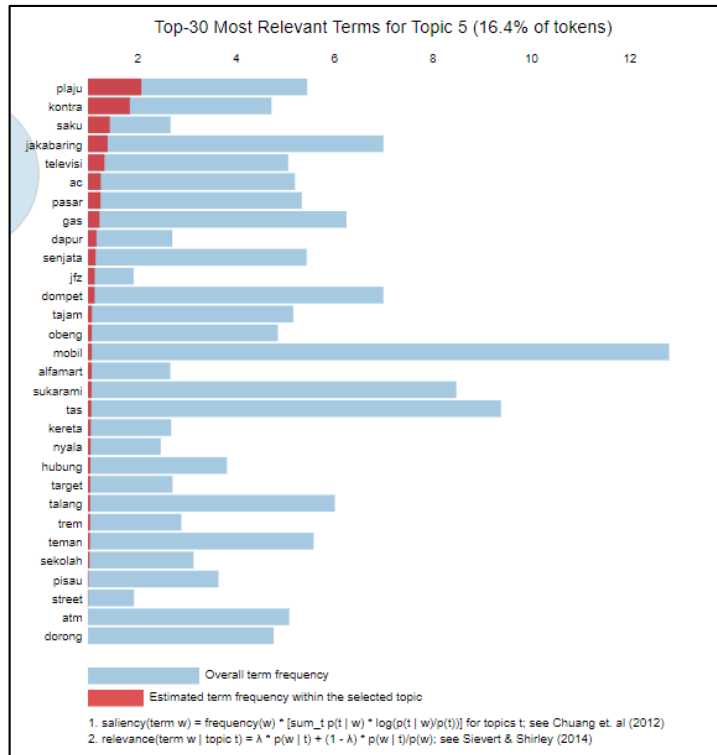
marginal topic distribution untuk topik 4 adalah sebesar 18.5%, yang artinya pentingnya topik terhadap dokumen sebesar 18.5%.

Tabel 5.36 Model Topik 4 Tahun 2022

Model Topik 4
$0.014 * \text{"tas"} + 0.012 * \text{"dompet"} + 0.011 * \text{"sukarami"} + 0.010 * \text{"atm"} + 0.009 * \text{"teman"} + 0.008 * \text{"kipas"} + 0.008 * \text{"gandus"} + 0.007 * \text{"ac"} + 0.006 * \text{"jakabaring"} + 0.006 * \text{"kemuning"}$

Tabel 5.36 menampilkan model pada topik 4, berdasarkan hasil model yang didapatkan, terlihat ada 4 nama daerah yaitu ‘sukarami’, ‘gandus’, ‘jakabaring’, dan ‘kemuning’ dari keempat daerah tersebut mengindikasikan 4 daerah yang banyak terjadi kasus pencurian. Sehingga, dapat peneliti simpulkan bahwa untuk topik 4 dari analisis LDA tahun 2022 membahas mengenai daerah tempat terjadinya pencurian. Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, kasus pertama terjadi di Kecamatan **Sukarami**, terdakwa merampas 1 unit *handphone* Oppo digengaman korban yang saat itu sedang bermain sepeda. Kasus kedua terjadi di Kelurahan **Gandus**, terdakwa mendatangi rumah korban yang saat itu sedang sepi, terdakwa mengambil pintu pagar milik korban dengan tujuan untuk dijual ke pengepul barang bekas.

Berikut ini merupakan visualisasi dengan menggunakan *pyLDAvis* dari topik ke-5 yang didapatkan:



Gambar 5.33 Visualisasi Topik 5 Tahun 2022

Pada **Gambar 5.33**, terlihat diagram batang yang berwarna merah memperlihatkan perkiraan *term frequency* dalam topik 5, diagram tersebut menunjukkan 30 terminologi yang paling relevan secara berurutan. Didapatkan nilai *marginal topic distribution* untuk topik 5 adalah sebesar 16.4%, yang artinya pentingnya topik terhadap dokumen sebesar 16.4%.

Tabel 5.37 Model Topik 5 Tahun 2022

Model Topik 5			
\0.011*	"plaju"	+	0.010*
			"kontra"
			+
			0.007*
			"saku"
			+
			0.007*
			"jakabaring"
			+
			0.007*
			"televisi"
			+
			0.006*
			"ac"
			+
			0.006*
			"pasar"
			+
			0.006*
			"gas"
			+
			0.006*
			"dapur"
			+
			0.006*
			"senjata"

Tabel 5.37 menampilkan model pada topik 5, berdasarkan hasil model yang didapatkan, dapat peneliti simpulkan bahwa untuk topik 5 dari analisis LDA tahun 2022 membahas mengenai jenis barang curian. Hal ini didasari kata ‘televisi’, ‘ac’, dan ‘gas’. Berikut contoh kasus pada Surat Dakwaan yang menyebutkan salah satu kata di atas, pada kasus ini terdakwa memasuki rumah korban dengan cara membuka pintu rumah menggunakan obeng yang telah dipersiapkan, selanjutnya terdakwa mengambil 1 unit **televisi** 32 inchi merk Sharp, lalu membawa **televisi** tersebut keluar dari rumah korban.

Tabel 5.38 Kesimpulan Topik Per Tahun

Tahun	Nilai <i>Coherence</i>	Jumlah Topik	Keterangan
2019	0.44983	3	Jenis barang curian, alat pendukung pencurian, metode pencurian.
2020	0.36465	6	Target pencurian, target pencurian, jenis pencurian, daerah tempat terjadinya pencurian, alat pendukung pencurian, daerah tempat terjadinya pencurian.
2021	0.40121	5	Alat pendukung pencurian, jenis pencurian, jenis barang curian, tempat terjadinya pencurian, target pencurian.
2022	0.42441	5	Target pencurian, tempat terjadinya pencurian, jenis barang curian, daerah tempat terjadinya pencurian, jenis barang curian.

Berdasarkan kesimpulan masing-masing topik di atas, maka didapatkan hasil perbandingan sebagai berikut:

Tabel 5.39 Hasil Perbandingan Topik Antar Tahun

Tahun	Hasil Perbandingan Topik
2019 & 2020	Berdasarkan hasil analisis tahun 2019, diketahui bahwa metode pencurian yang banyak dilakukan oleh para terdakwa adalah dengan cara menjambret, kemudian, jenis barang curian yang teridentifikasi berupa dompet serta uang tunai, hal ini bersesuaian dengan metode pencurian yang dilakukan terdakwa. Berdasarkan hasil perbandingan, terdapat karakteristik baru yang teridentifikasi terkait kasus tindak pidana pencurian berdasarkan topik tahun 2020 yang didapatkan, yaitu mengenai target pencurian, jenis pencurian, dan daerah tempat terjadinya pencurian. Berdasarkan topik yang didapatkan, diketahui bahwa target pencurian berupa toko-toko serta perkantoran. Berdasarkan topik yang didapatkan dari analisis tahun 2020, diketahui bahwa jenis pencurian yang

	<p>marak terjadi pada tahun 2020 adalah pencurian dengan kekerasan. Serta, pencurian banyak terjadi di daerah Plaju, Kalidoni, Kemuning, Bukit, dan Jakabaring.</p>
2020 & 2021	<p>Berdasarkan hasil perbandingan, didapatkan bahwa pada tahun 2021 terdapat karakteristik baru yang teridentifikasi terkait kasus tindak pidana pencurian, berupa jenis barang curian dan tempat terjadinya pencurian. Diketahui bahwa jenis pencurian yang marak terjadi pada tahun 2021 adalah curanmor, baik curanmor R2 maupun curanmor-R4. Pencurian dapat terjadi dimanapun dan kapanpun, sebagai contoh, pada tahun 2021 pencurian bahkan terjadi di sebuah rumah ibadah, dimana pelaku membobol suatu masjid dan mengambil barang-barang di dalamnya.</p>
2021 & 2022	<p>Berdasarkan hasil perbandingan, diketahui bahwa tidak terdapat karakteristik baru yang teridentifikasi pada tahun 2022 dengan tahun sebelumnya. Namun, berdasarkan model yang didapatkan diketahui bahwa terdapat perbedaan daerah tempat terjadinya pencurian, pada tahun 2022 pencurian banyak terjadi di daerah Sukarami, Gandus, Jakabaring, dan Kemuning.</p>

BAB VI

PENUTUP

6.1. Kesimpulan

Berdasarkan hasil analisis dan pembahasan, maka didapatkan kesimpulan sebagai berikut.

3. Diketahui, waktu dengan jumlah kasus pencurian paling banyak adalah pada malam hari, hal ini disebabkan pada waktu tersebut para korban tidak memiliki upaya untuk mencegah tindak pidana pencurian. Jenis tindak pidana pencurian dengan jumlah kasus tertinggi adalah tindak pidana pencurian dengan pemberatan. Tahun 2022 memiliki rata-rata lama hukuman (bulan) yang paling tinggi jika dibandingkan dengan tahun lainnya.
4. Dengan menggunakan *Topic Modeling* dengan metode *Latent Dirichlet Allocation* (LDA) jumlah topik yang terbentuk pada tahun 2019, 2020, 2021, dan 2022 secara berturut-turut adalah 3 topik, 6 topik, 5 topik, dan 5 topik. Karakteristik yang teridentifikasi berdasarkan analisis masing-masing tahun adalah sebagai berikut:
 - a. Tahun 2019: Jenis barang curian, alat pendukung pencurian, metode pencurian.
 - b. Tahun 2020: Target pencurian, target pencurian, jenis pencurian, daerah tempat terjadinya pencurian, alat pendukung pencurian, daerah tempat terjadinya pencurian.
 - c. Tahun 2021: Alat pendukung pencurian, jenis pencurian, jenis barang curian, tempat terjadinya pencurian, target pencurian.
 - d. Tahun 2022: target pencurian, tempat terjadinya pencurian, jenis barang curian, daerah tempat terjadinya pencurian, jenis barang curian.
5. Berdasarkan hasil perbandingan topik antar tahun, didapatkan *insight* terkait karakteristik kasus tindak pidana pencurian sebagai berikut:
 - a. Tahun 2019 & 2020: karakteristik baru yang teridentifikasi berupa target pencurian, jenis pencurian, dan daerah tempat terjadinya pencurian.
 - b. Tahun 2020 & 2021: karakteristik baru yang teridentifikasi berupa jenis barang curian, dan tempat terjadinya pencurian.

- c. Tahun 2021 & 2022: tidak terdapat perbedaan hasil yang teridentifikasi pada tahun 2022 dengan tahun sebelumnya terkait karakteristik kasus tindak pidana pencurian. Namun, diketahui terdapat perbedaan daerah tempat terjadinya pencurian yang teridentifikasi berdasarkan model yang didapatkan. Pada tahun 2022, pencurian banyak terjadi di daerah Sukarame, Gandus, Jakabaring, dan Kemuning.

6.2. Saran

Berikut saran yang diberikan berdasarkan hasil penelitian ini:

1. Pada penelitian ini hanya menggunakan metode LDA untuk melakukan analisis *Topic Modeling*, untuk penelitian selanjutnya dapat mengembangkan penelitian dengan menggunakan metode lain.
2. Untuk analisis *Topic Modeling* hanya menggunakan dokumen Surat Dakwaan, untuk penelitian selanjutnya dapat menambahkan data lainnya seperti data Putusan untuk kasus tindak pidana pencurian.
3. Masyarakat dapat lebih waspada dengan cara selalu mengunci pintu dan pagar ketika hendak bepergian, menambahkan kunci tambahan pada kendaraan dan rumah untuk meningkatkan keamanan.

DAFTAR PUSTAKA

- Adi, S. (2018). PERANCANGAN KLASIFIKASI TWEET BERDASARKAN SENTIMEN DAN FITUR CALON GUBERNUR DKI JAKARTA 2017. *Journal of Informatic Pelita Nusantara*.
- Apriani, Zakiyudin, H., & Marzuki, K. (2021). Penerapan Algoritma *Cosine Similarity* dan Pembobotan *TF-IDF System* Penerimaan Mahasiswa Baru pada Kampus Swasta. *Jurnal Bumigora Information Technology*.
- Ardianti, V. I., & Karmila, S. (2022). METODE LATENT DIRICHLET ALLOCATION UNTUK MENENTUKAN TOPIK TEKS SUATU BERITA. *Jurnal Informatika & Komputasi*.
- Arif, M. (2014). TINJAUAN KRIMINOLOGIS TERHADAP KEJAHATAN PERDAGANGAN ANAK. *Jurnal Ilmu Hukum Legal Opinion*.
- Cikania, R. N. (2021). Implementasi Algoritma *Naive Bayes Classifier* dan *Support Vector Machine* Pada Klasifikasi Sentimen *Review* Layanan Telemedicine Halodoc. *Jambura Journal of Probability and Statistics*.
- Dikiyanti, T. D., Rukmini, A. M., & Irawan, M. I. (2021). *Sentiment Analysis and Topic Modeling of BPJS Kesehatan Based on Twitter Crawling Data Using Indonesian Sentiment Lexicon and Latent Dirichlet Allocation Algorithm*. *Journal of Physics: Conference Series*.
- Fauzi, M. Y., & Jainah, Z. O. (2022). ANALISIS PERTANGGUNGJAWABAN PIDANA TERHADAP PELAKU YANG MELAKUKAN TINDAK PIDANA PERCOBAAN PENCURIAN DENGAN PEMBERATAN. *Jurnal Suara Keadilan*.
- Fudholi, D. H., & Ridhwanullah, D. (2022). Pemodelan Topik pada Cuitan tentang Penyakit Tropis di Indonesia dengan Metode *Latent Dirichlet Allocation*. *Jurnal Ilmiah Sinus (JIS)*.
- Gangadharan, V., & Gupta, D. (2020). *Recognizing Named Entities in Agriculture Documents using LDA based Topic Modeling Techniques*. *Third International Conference on Computing and Network Communications*.
- George, S., & Srividhya, V. (2020). *Comparison of LDA and NMF Topic Modeling Techniques for Restaurant Reviews*. *Indian Journal of Natural Science*.

- Gojali, H. A., Rosid, M. A., & Sumarno. (2020). Klasifikasi Keluhan Mahasiswa dengan Metode *Naive Bayes* dan Sastrawi. *Journal of Informatics, Network, and Computer Science*.
- Gunawan, B., Pratiwi, H. S., & Pratama, E. E. (2018). Sistem Analisis Sentimen pada Ulasan Produk Menggunakan Metode *Naive Bayes*. *Jurnal Edukasi Dan Penelitian Informatika*.
- Hutapea, M. G., Siburian, K., & Sinaga, J. (2020). PERTANGGUNGJAWABAN PIDANA TERHADAP PELAKU YANG MELAKUKAN PENGANGKUTAN BAHAN BAKAR MINYAK (BBM) TANPA IZIN USAHA PENGANGKUTAN. *PATIK: Jurnal Hukum*.
- Hutasuhut, S. H. (2022). PERANAN STATISTIKA DALAM PENELITIAN PENDIDIKAN MATEMATIKA. *MES: Journal of Mathematics Education and Science*.
- Ikegami, A., & Darmawan, D. M. B. A. (2022). Analisis Sentimen dan Pemodelan Topik Ulasan Aplikasi Noice Menggunakan *XGBoost* dan LDA. *Jurnal Nasional Teknologi Informasi Dan Aplikasinya*.
- Jumeilah, S. F. (2017). Penerapan *Support Vector Machine* (SVM) untuk Pengkategorian Penelitian. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*.
- Laxmi, M. D., Indriati, & Fauzi, M. A. (2019). *Query Expansion* Pada Sistem Temu Kembali Informasi Berbahasa Indonesia dengan Metode Pembobotan TF-IDF dan Algoritme *Cosine Similarity* Berbasis Wordnet. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*.
- Mansyur, J., Christanti, V., & Sutrisno, T. (2019). ANALISIS PENDAPATAN PUBLIK TERHADAP PUBLIC FIGURE DENGAN MENGGUNAKAN METODE NAIVE BAYES. *Jurnal Ilmu Komputer Dan Sistem Informasi*.
- Mardi, Y. (2017). Data Mining: Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika*.
- McNaught, C., & Lam, P. (2010). Using Wordle as a Supplementary Research Tool. *The Qualitative Report*.
- Nasution, K. H., Widodo, & Adhi, bambang P. (2021). SISTEM DETEKSI TOPIK POLITIK PADA TWITTER MENGGUNAKAN ALGORITMA LATENT

DIRICHLET ALLOCATION. *Jurnal PINTER*.

- Natalia, C., Suprata, F., Surbakti, F. P. S., & Clarence, S. (2021). Penentuan Standar Spesifikasi Kerja di Cafe Berdasarkan Big Data dengan Metode LDA dan AHP. *Jurnal Rekayasa Sistem Industri*.
- Naury, C., Fudholi, D. H., & Hidayatullah, A. F. (2021). Topic Modeling pada Sentimen Terhadap Headline Berita Online Berbahasa Indonesia Menggunakan LDA dan LSTM. *Jurnal Media Informatika Budidarma*.
- Novianti, Amin, M., & Kifti, W. M. (2021). METODE EXPONENTIAL SMOOTHING PADA PERAMALAN TINGKAT KRIMINALITAS PENCURIAN SEPEDA MOTOR. *Journal of Computer*.
- Pranata, M., Anggraini, D., Makbuloh, D., & Rinaldi, A. (2020). PREDIKSI PENCURIAN SEPEDA MOTOR MENGGUNAKAN MODEL TIME SERIES (STUDI KASUS: POLRES KOTABUMI LAMPUNG UTARA). *Jurnal Ilmu Matematika Dan Terapan*.
- Putri, N. W. N. S. T., & Giri, R. Ri. W. (2020). ANALISIS TOPIC MODELING UNTUK MENGIDENTIFIKASI TOPIK PEMBICARAAN PADA MEDIA SOSIAL BANG BNI. *Jurnal Mitra Manajemen (JMM Online)*.
- Rahim, A. (2019). Internalisasi Nilai Sipakatau, Sipakalebbi, Sipakainge' Dalam Upaya Pencegaham Tindak Pidana Korupsi. *Jurnal Al-Himayah*.
- Rahmawati, T. (2014). APLIKASI PRINCIPAL COMPONENT ANALYSIS (PCA) UNTUK MEREDUKSI FAKTOR-FAKTOR YANG BERPENGARUH DALAM PERAMALAN KONSUMSI LISTRIK. *Journal of Technology and Informatics*.
- Rashif, F., Nirvana, G. I. P., Noor, M. A., & Rakhmawati, N. A. (2021). Implementasi LDA untuk Pengelompokan Topik Cuitan Akun Bot Twitter Bertagar #Covid-19. *Cogito Smart Journal*.
- Risnantoyo, R., Nugroho, A., & Mandara, K. (2020). Sentiment Analysis on Corona Virus Pandemic Using Machine Learning Algorithm. *JITE (Journal of Informatics and Telecommunication Engineering)*.
- Rumahorbo, E. L. D. ., Sihombing, L. N., & Sitio, H. (2022). Pengaruh Media Kartu Huruf terhadap Prestasi Belajar Siswa dalam Mata Pelajaran Bahasa Indonesia di Kelas I SD Negeri 091621 Perdagangan. *Jurnal Pendidikan Dan Konseling*.

- Salsabila, K. D. A., & Trianasari, N. (2021). Analisis Persepsi Produk Kosmetik Menggunakan Metode *Sentiment Analysis* dan *Topic Modeling* (Studi Kasus: Laneige Water Sleeping Mask). *Jurnal Teknologi Dan Manajemen Informatika*.
- Saputra, R. P. (2019). PERKEMBANGAN TINDAK PIDANA PENCURIAN DI INDONESIA. *Jurnal Pahlawan*.
- Sari, I. (2020). PERBUATAN MELAWAN HUKUM (PHM) DALAM HUKUM PIDANA DAN HUKUM PERDATA. *Jurnal Ilmiah Hukum Dirgantara*.
- Setijohatmo, U. T., Rachmat, S., Susilawati, T., & Rahman, Y. (2020). Analisis Metoda *Latent Dirichlet Allocation* untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik. *Prosiding The 11th Industrial Research Workhop and National Seminar*.
- Siregar, R. R. A., Sinaga, F. A., & Arianto, R. (2017). APLIKASI PENENTUAN DOSEN PENGUJI SKRIPSI MENGGUNAKAN METODE TF-IDF DAN VECTOR SPACE MODEL. *Journal of Computer and Information System*.
- Suari, N. M. E. P., Widyantara, I. M. M., & Karma, N. M. S. (2020). KEDUDUKAN DAN PERLINDUNGAN SAKSI MAHKOTA DALAM TINDAK PIDANA PENCURIAN DENGAN KEKERASAN. *Jurnal Interpretasi Hukum*.
- Suryadi, U. T., & Supriatna, Y. (2019). SISTEM CLUSTERING TINDAK KEJAHATAN PENCURIAN DI WILAYAH JAWA BARAT MENGGUNAKAN ALGORITMA K-MEANS. *Jurnal Teknologi Informasi Dan Komunikasi*.
- Susanto, H., & Sudiyatno. (2014). DATA MINING UNTUK MEMPREDIKSI PRESTASI SISWA BERDASARKAN SOSIAL EKONOMI, MOTIVASI, KEDISIPLINAN DAN PRESTASI MASA LALU. *Jurnal Pendidikan Vokasi*.
- Syaifuddin, A., Harianto, R. A., & Santoso, dan J. (2020). Analisis Trending Topik untuk Percakapan Media Sosial dengan Menggunakan *Topic Modelling* Berbasis Algoritme LDA. *Journal of Intelligent Systems and Computation*.
- Tresnasari, N. A., Adji, T. B., & Permanasari, A. E. (2020). *Social-Child-Case Document Clustering based on Topic Modeling using Latent Dirichlet Allocation*. *Indonesian Journal of Computing and Cybernetics Systems*.

- Utami, E., & Suparyati. (2022). PENGAMATAN TREN ULASAN HOTEL MENGGUNAKAN PEMODELAN TOPIK BERBASIS LATENT DIRICHLET ALLOCATION. *JIKO (Jurnal Informatika Dan Komputer)*.
- Utami, M. T., Tulili, T. R., & Topadang, A. (2017). Implementasi Metode *City Block Distance* pada Identifikasi Citra Tanda Tangan. *Jurnal Teknologi Terpadu*.
- Utomo, P. E. P., Manaar, Khaira, U., & Suratno, T. (2019). ANALISIS SENTIMEN ONLINE REVIEW PENGGUNA BUKALAPAK MENGGUNAKAN METODE ALGORITMA TF-DF. *Jurnal Sains Dan Sistem Informasi*.
- Zazin, M., Sulaksono, D. H., Yuliasuti, G. E., & Prabiantissa, C. N. (2022). *Implementasi IoT pada Sistem Surveillance Camera Via Telegram*. *Jurnal Teknologi Dan Manajemen*.

