

**IMPLEMENTASI *LATENT DIRICHLET ALLOCATION*  
TERHADAP DATA KASUS TINDAK PIDANA**

(Studi Kasus : Surat Dakwaan Kasus Tindak Pidana di Pengadilan Negeri  
Yogyakarta Pada Tahun 2019-2021)

**TUGAS AKHIR**

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



Disusun Oleh:

Dela Gustiara

19611066

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

**HALAMAN PERSETUJUAN PEMBIMBING  
TUGAS AKHIR**

Judul : Implementasi *Latent Dirichlet Allocation*  
Terhadap Data Kasus Tindak Pidana  
(Studi Kasus : Surat Dakwaan Kasus Tindak  
Pidana di Pengadilan Negeri Yogyakarta Pada  
Tahun 2019-2021)

Nama Mahasiswa : Dela Gustiara

NIM : 19611066

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

Yogyakarta, .....<sup>07</sup>/<sub>03</sub> - 2023

Pembimbing

  
The signature is written in black ink over a background of yellow Arabic calligraphy. The calligraphy includes the words 'UNIVERSITAS ISLAM INDONESIA' and 'الجمعة البعيدة انبسطت'.

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

**HALAMAN PENGESAHAN**  
**TUGAS AKHIR**

**IMPLEMENTASI *LATENT DIRICHLET ALLOCATION* TERHADAP  
DATA KASUS TINDAK PIDANA**

(Studi Kasus : Surat Dakwaan Kasus Tindak Pidana di Pengadilan Negeri  
Yogyakarta Pada Tahun 2019-2021)

Nama Mahasiswa : Dela Gustiara

NIM : 19611066

TUGAS AKHIR INI TELAH DIUJIKAN

PADA TANGGAL : 6 April 2023

Nama Penguji

Tanda Tangan

1. Dr. Raden Bagus Fajriya Hakim, S.Si., M.Si
2. Muhammad Hasan Sidiq Kurniawan, 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*

*Alhamdulillah* puji syukur kita panjatkan atas kehadiran Allah SWT yang telah melimpahkan rahmat, anugerah dan hidayah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “Implementasi *Latent Dirichlet Allocation* Terhadap Data Kasus Tindak Pidana (Studi Kasus : Surat Dakwaan Kasus Tindak Pidana di Pengadilan Negeri Yogyakarta pada Tahun 2019-2021)” ini sebagaimana mestinya. Shalawat serta salam tidak lupa di haturkan kepada Nabi Muhammad SAW, keluarga, dan para sahabatnya yang telah menuntun ke jalan yang penuh berkah ini.

Tugas Akhir ini disusun guna memenuhi persyaratan meraih gelar sarjana Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia. Penulis menyadari dalam proses penyusunan Tugas Akhir ini banyak memperoleh bantuan dan dukungan dari berbagai pihak, baik berupa saran, kritik maupun bantuan lainnya. Maka dari itu, pada kesempatan ini penulis 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 Program Studi Statistika Universitas Islam Indonesia.
3. Ibu Ayundyah Kesumawati, S.Si., M.Si., selaku Dosen Pembimbing Tugas Akhir yang telah memberikan ilmu, motivasi, bimbingan dan arahan selama penyusunan Tugas Akhir ini.
4. Seluruh Dosen Pengajar dan Staff Program Studi Statistika yang telah banyak memberikan ilmu serta bimbingannya kepada penulis selama perkuliahan berlangsung.
5. Bapak Suhardi dan Ibu Nuryantina selaku orangtua penulis yang telah merawat, menjaga dengan penuh rasa kasih sayang dan selalu memberikan dukungan kepada penulis selama ini.

6. Ibu Nuryantini, Bapak Erwin Asmadi dan seluruh keluarga besar yang telah banyak memberikan bantuan, dukungan dan menyayangi penulis selama ini.
7. Kepada Mba Rara yang telah banyak memberikan bantuan, semangat, motivasi dan bimbinganya selama perkuliahan dan penyelesaian Tugas Akhir ini.
8. Teman-teman penulis, Meutya Rahmadini, Anisa Dwi Mulyaningsih, Nadhilah Fadia Haya, Anis Fitriana As Shidiqi, Nindala Divaoni Ayudhia yang selalu memberikan semangat dan menemani penulis selama perkuliahan baik suka maupun duka.
9. Teman-teman penulis di grup “Alhamdulillah” (Rahmi, Shabira, Mifta, Aulia, Yosi), Naelun, Rizqia dan teman KKN yang selalu berbagi ilmu, berbagi cerita dan pengalamannya selama perkuliahan.
10. Semua teman-teman statistika Angkatan 2019 yang selalu memberikan semangat dan pengalamannya kepada penulis.
11. Semua pihak yang telah membantu yang tidak dapat di sebutkan satu persatu, terima kasih atas dukungan dan semangat yang telah diberikan.
12. Terimakasih kepada Dela Gustiara yang telah kuat dan mampu bertahan sampai sejauh ini.

Penulis menyadari sepenuhnya bahwa Tugas Akhir ini masih jauh dari kata sempurna. Oleh karena itu segala kritik dan saran diharapkan untuk perbaikan penelitian ini menjadi lebih baik. Semoga Tugas Akhir ini dapat bermanfaat bagi penulis, pembaca, maupun peneliti lain di masa yang akan datang. Demikianlah yang dapat disampaikan, semoga Allah SWT senantiasa melimpahkan rahmat dan ridho-Nya kepada semua pihak yang telah membantu penulis.

***Wassalamu'alaikum Wr. Wb***

Yogyakarta, 7 Maret 2023

Dela Gustiara

## DAFTAR ISI

HALAMAN PENGESAHAN TUGAS AKHIR .....	iii
KATA PENGANTAR .....	iv
DAFTAR ISI .....	vi
DAFTAR TABEL .....	viii
DAFTAR GAMBAR .....	ix
DAFTAR LAMPIRAN .....	x
PERNYATAAN .....	xi
INTISARI .....	xii
ABSTRACT .....	xiii
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang Masalah .....	1
1.2. Rumusan Masalah .....	3
1.3. Batasan Masalah .....	3
1.4. Tujuan Penelitian .....	3
1.5. Manfaat Penelitian .....	4
BAB II TINJAUAN PUSTAKA .....	5
BAB III LANDASAN TEORI .....	9
3.1. Pidana .....	9
3.1.1 Jenis-Jenis Pidana .....	9
3.1.2 Jenis-Jenis Acara Pemeriksaan Pidana .....	11
3.2. Surat Dakwaan .....	12
3.3. Statistika Deskriptif .....	12
3.4. <i>Machine Learning</i> .....	13
3.5. <i>Text Mining</i> .....	13
3.6. <i>Text Processing</i> .....	14
3.7. <i>Topic Modelling</i> .....	15
3.8. <i>Term Frequency-Inverse Document Frequency (TF-IDF)</i> .....	16
3.9. <i>Latent Dirichlet Allocation (LDA)</i> .....	18
3.10. <i>Topic Coherence</i> .....	20
BAB IV METODOLOGI PENELITIAN .....	22
4.1. Populasi Penelitian .....	22
4.2. Jenis Data dan Sumber Data .....	22
4.3. Variabel penelitian .....	22
4.4. Metode Analisis Data .....	23
4.5. Tahapan Penelitian .....	23
BAB V HASIL DAN PEMBAHASAN .....	26
5.1. Analisis Deskriptif dan Uji Independensi .....	26
5.1.1 Analisis Deskriptif .....	26
5.2. <i>Preprocessing</i> .....	28
5.2.1 <i>Case Folding</i> .....	29
5.2.2 <i>Remove Punctuation dan Numbers</i> .....	30
5.2.3 <i>Stopword</i> .....	32
5.2.4 <i>Stemming</i> .....	33
5.2.5 <i>Tokenizing</i> .....	34
5.3. <i>Term Frequency-Invers Document Frequency (TF-IDF)</i> .....	35

5.4.	Hasil <i>Topic Modelling</i> .....	38
5.4.1	Narkotika .....	38
5.4.2	Pencurian .....	45
5.4.3	Kesehatan .....	51
6.1.	Kesimpulan .....	58
6.2.	Saran.....	59
	DAFTAR PUSTAKA.....	60
	LAMPIRAN .....	64

## DAFTAR TABEL

<b>Tabel 2.1</b> Deskripsi Penelitian Terdahulu .....	7
<b>Tabel 4.1</b> Variabel Penelitian .....	22
<b>Tabel 5.1</b> Data Surat Dakwaan Pidana.....	28
<b>Tabel 5.2</b> Hasil <i>Case Folding</i> .....	29
<b>Tabel 5.3</b> Hasil <i>Removing Punctuation dan Numbers</i> .....	31
<b>Tabel 5.4</b> Hasil <i>Stopword</i> .....	32
<b>Tabel 5.5</b> Hasil <i>Stemming</i> .....	34
<b>Tabel 5.6</b> Hasil <i>Tokenizing</i> .....	35
<b>Tabel 5.7</b> Hasil Pembobotan <i>TF-IDF</i> dengan <i>Phyton</i> .....	36
<b>Tabel 5.8</b> Hasil Pembobotan <i>TF</i> dengan <i>Phyton</i> .....	36
<b>Tabel 5.9</b> Hasil Pembobotan <i>DF</i> Perhitungan Manual.....	37
<b>Tabel 5.10</b> Hasil Pembobotan <i>TD-IDF</i> Perhitungan Manual.....	37
<b>Tabel 5.11</b> Nilai Koherensi Topik Narkotika.....	39
<b>Tabel 5.12</b> Hasil <i>Pricipal Component</i> .....	40
<b>Tabel 5.13</b> Hasil Pemodelan 3 Topik Narkotika .....	44
<b>Tabel 5.14</b> Kesimpulan Pemodelan 3 Topik Narkotika .....	44
<b>Tabel 5.15</b> Nilai Koherensi Topik Pencurian.....	45
<b>Tabel 5.16</b> Hasil <i>Pricipal Component</i> .....	47
<b>Tabel 5.17</b> Hasil Pemodelan 3 Topik Pencurian .....	50
<b>Tabel 5.18</b> Kesimpulan Pemodelan 3 Topik Pencurian .....	51
<b>Tabel 5.19</b> Nilai Koherensi Topik Kesehatan .....	52
<b>Tabel 5.20</b> Hasil <i>Pricipal Component</i> .....	53
<b>Tabel 5.21</b> Hasil Pemodelan 2 Topik Kesehatan .....	56
<b>Tabel 5.22</b> Kesimpulan Pemodelan 2 Topik Kesehatan .....	57



## DAFTAR GAMBAR

<b>Gambar 3.1</b> Proses <i>Text Mining</i> .....	14
<b>Gambar 3.2</b> Konsep <i>Topic Modelling</i> .....	16
<b>Gambar 3.3</b> Representasi Model Grafis <i>LDA</i> .....	18
<b>Gambar 4.1</b> Diagram Alir Penelitian.....	24
<b>Gambar 5.1</b> Diagram Publikasi Kasus Tindak Pidana .....	26
<b>Gambar 5.2</b> Diagram Klasifikasi Perkara Kasus Tindak Pidana.....	27
<b>Gambar 5.3</b> Diagram Denda dan Hukuman Kasus Tindak Pidana .....	27
<b>Gambar 5.4</b> Grafik <i>Coherence Score</i> Narkotika .....	38
<b>Gambar 5.5</b> Visualisasi 3 Topik Perkara Narkotika.....	39
<b>Gambar 5.6</b> <i>Word Cloud</i> Perkara Narkotika .....	41
<b>Gambar 5.7</b> Visualisasi Topik 1 .....	42
<b>Gambar 5.8</b> Visualisasi Topik 2 .....	42
<b>Gambar 5.9</b> Visualisasi Topik 3 .....	43
<b>Gambar 5.10</b> Grafik <i>Coherence Score</i> Pencurian .....	45
<b>Gambar 5.11</b> Visualisasi 3 Topik Perkara Pencurian.....	46
<b>Gambar 5.12</b> <i>Word Cloud</i> Perkara Pencurian .....	47
<b>Gambar 5.13</b> Visualisasi Topik 1 .....	48
<b>Gambar 5.14</b> Visualisasi Topik 2 .....	49
<b>Gambar 5.15</b> Visualisasi Topik 3 .....	49
<b>Gambar 5.16</b> Grafik <i>Coherence Score</i> Kesehatan.....	51
<b>Gambar 5.17</b> Visualisasi 2 Topik Perkara Kesehatan .....	53
<b>Gambar 5.18</b> <i>Word Cloud</i> Perkara Kesehatan.....	54
<b>Gambar 5.19</b> Visualisasi Topik 1 .....	55
<b>Gambar 5.20</b> Visualisasi Topik 2 .....	55

## DAFTAR LAMPIRAN

<b>Lampiran 1</b> Data Kasus Tindak Pidana.....	64
---	----

## 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, 7 Maret 2023



Dela Gustiara

## INTISARI

### IMPLEMENTASI *LATENT DIRICHLET ALLOCATION* TERHADAP DATA KASUS TINDAK PIDANA

(Studi Kasus : Surat Dakwaan Kasus Tindak Pidana di Pengadilan Negeri  
Yogyakarta Pada Tahun 2019-2021)

Dela Gustiara

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

Kriminalitas merupakan suatu tindakan pidana yang melanggar hukum karena berkaitan dengan mengambil hak milik orang lain dan dapat menimbulkan suatu kerugian, ketidaksenangan, dan ketidaklayakan dalam kehidupan bermasyarakat. Kemajuan ilmu pengetahuan dan teknologi telah memberikan peluang semakin tinggi dan meningkat kualitas dari tindak kejahatan. Menurut laporan Kepolisian Republik Indonesia berdasarkan data Pusat Informasi Kriminal Nasional terdapat 274.988 kasus tindak pidana pada tahun 2021, 247.780 kasus tindak pidana pada tahun 2020, dan 178.207 kasus tindak pidana pada tahun 2019. Daerah Istimewa Yogyakarta merupakan salah satu daerah yang memiliki tingkat kejahatan tergolong tinggi. Menurut kepolisian tercatat bahwa pada tahun 2021 kasus kejahatan di Daerah Istimewa Yogyakarta mengalami kenaikan dibandingkan tahun 2020 dan 2019. Hal tersebut menjadi perhatian tersendiri bagi pemerintah dan masyarakat untuk menekan tingginya kasus kejahatan. Adanya surat dakwaan yang menguraikan penyebab dan faktor terjadinya tindak pidana dapat dimanfaatkan untuk memperoleh topik utama dari kumpulan dokumen surat dakwaan. Maka penelitian ini dilakukan menggunakan *topic modelling* dengan tujuan untuk mengetahui topik kasus tindak pidana yang sering terjadi dan dibahas dari klasifikasi perkara narkoba, pencurian dan kesehatan di Daerah Istimewa Yogyakarta. Dengan menggunakan *topic modelling* yaitu *Latent Dirichlet Allocation (LDA)* berdasarkan surat dakwaan kasus tindak pidana di Pengadilan Negeri Yogyakarta pada tahun 2019-2021 yang mempunyai 971 kasus diperoleh topik yang sering dibahas untuk 3 klasifikasi perkara tertinggi yaitu narkoba terdapat 3 topik, pencurian terdapat 3 topik dan kesehatan terdapat 2 topik.

**Kata Kunci** : Kejahatan, Tindak Pidana, *Topic Modelling*, *Latent Dirichlet Allocation*.

## ABSTRACT

### **IMPLEMENTATION OF LATENT DIRICHLET ALLOCATION ON CRIMINAL CASES DATA**

*(Case Study : Indictment Of Criminal Cases at The Yogyakarta District  
Court in 2019-2021)*

Dela Gustiara

Department of Statistics, Faculty of Mathematics and Natural Sciences  
Universitas Islam Indonesia

*Crime is a criminal act that violates the law because it is related to taking the property of others and can cause harm, displeasure, and injustice in social life. Advances in science and technology have provided opportunities for higher and higher quality crimes. According to a report by the Indonesian National Police based on data from the National Criminal Information Center, there were 274.988 criminal cases in 2021, 247.780 criminal cases in 2020, and 178.207 criminal cases in 2019. The Special Region of Yogyakarta is one of the regions that has a high crime rate. According to the police, it was recorded that in 2021 crime cases in the Special Region of Yogyakarta increased compared to 2020 and 2019. This is a concern for the government and the community to reduce the high crime rate. The existence of an indictment that describes the causes and factors of a criminal offense can be utilized to obtain the main topics from a collection of indictment documents. So this research was conducted using topic modeling with the aim of knowing the topic of criminal cases that often occur and are discussed from the classification of narcotics, theft and health cases in the Special Region of Yogyakarta. By using topic modeling, namely Latent Dirichlet Allocation (LDA) based on the indictment of criminal cases in the Yogyakarta District Court in 2019-2021 which has 971 cases, it is obtained that the topics that are often discussed for the 3 highest case classifications are narcotics there are 3 topics, theft there are 3 topics and health there are 2 topics.*

**Keywords:** *Crime, Criminal Act, Topic Modelling, Latent Dirichlet Allocation.*

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang Masalah**

Era sekarang kejahatan seperti hal yang sudah biasa di masyarakat, ketika seseorang tidak perlu berpikir panjang untuk melakukan tindak kejahatan dan para pelaku tidak lagi memikirkan konsekuensi yang terjadi dari perbuatannya tersebut. Perilaku tindak kejahatan dapat dilakukan oleh siapapun, pria maupun wanita dari usia anak-anak, dewasa hingga usia lanjut. Kejahatan dapat dilakukan secara sadar yaitu dengan dipikirkan, direncanakan dan diarahkan terlebih dahulu untuk suatu maksud tertentu dalam keadaan yang benar-benar sadar. Selain itu untuk kejahatan yang dilakukan setengah sadar yaitu dipengaruhi oleh dorongan atau obsesi yang kuat dalam diri seseorang. Serta, suatu kejahatan dapat juga dilakukan dalam keadaan yang tidak sadar yaitu secara terpaksa (Septiani et al., 2020).

Kejahatan atau kriminalitas merupakan suatu tindakan pidana yang melanggar hukum karena berkaitan dengan mengambil hak milik orang lain dan dapat menimbulkan suatu kerugian, ketidaksenangan, dan ketidaklayakan dalam kehidupan bermasyarakat. Berbagai macam kerugian yang dapat ditimbulkan, baik dari segi materil yaitu ekonomis, maupun dari segi immateril seperti hilangnya rasa aman, nyaman dan ketentraman dalam kehidupan bermasyarakat. Kejahatan sangat berpengaruh dalam kehidupan bermasyarakat dan banyak faktor yang mempengaruhi seseorang melakukan kejahatan yaitu faktor personal, faktor sosial dan faktor situasional (Weda, 1996).

Kemajuan ilmu pengetahuan dan teknologi yang tinggi memberi peluang kejahatan makin tinggi dan meningkat kualitasnya, termasuk pelanggaran pidana yang makin bervariasi. Menurut Kepolisian Republik Indonesia berdasarkan data Pusat Informasi Kriminal Nasional (Pusiknas) terdapat 274.988 kasus tindak pidana sepanjang tahun 2021 di Indonesia. Jumlah tersebut mengalami kenaikan dibandingkan tahun 2019 dan 2020. Sepanjang tahun 2020 terdapat 247.780 kasus, sedangkan sepanjang 2019 terdapat 178.207 kasus (Widi, 2022).

Daerah Istimewa Yogyakarta merupakan salah satu daerah yang memiliki tingkat kejahatan tergolong tinggi. Menurut Kepolisian Daerah Istimewa

Yogyakarta terdapat kenaikan penanganan kasus kejahatan yang terjadi sepanjang 2021. Berdasarkan catatan kepolisian, total kejahatan yang terjadi pada tahun 2021 mengalami kenaikan sebanyak 12,24% dibandingkan tahun 2020. Sementara mengacu pada data yang sama, tingkat kerawanan kejahatan pada area tertentu sepanjang 2021 juga meningkat yaitu 12,59% per 10.000 penduduk. Untuk kasus kejahatan terbanyak yaitu narkoba dengan jumlah 527 kasus, pencurian dengan jumlah 339 kasus dan kesehatan dengan jumlah 42 kasus (Leon, 2021). Hal ini tentunya menjadi perhatian tersendiri bagi pemerintah dan masyarakat setempat untuk terus menekan tingginya kasus kejahatan.

Upaya yang dapat dilakukan untuk penanggulangan masalah kejahatan dengan memberikan sanksi berupa pidana yang jenisnya tertuang dalam KUHP Pasal 10 (Timbuleng, 2020). Untuk mengadili individu/kelompok yang melakukan kejahatan terhadap hukum diadili melalui pengadilan negeri dimana kejahatan tersebut dilakukan. Jika kejahatan terjadi di daerah Kota Yogyakarta maka akan diadili melalui Pengadilan Negeri Yogyakarta. Tugas Pengadilan Negeri Yogyakarta adalah sebagai pengadilan tingkat pertama yang bertugas dan berwenang menerima, memeriksa, serta memutus perkara yang masuk di tingkat pertama (Yogyakarta, 2019). Dalam hal ini, penyebab dan faktor terjadi tindak pidana diuraikan menggunakan surat dakwaan. Surat dakwaan dijadikan dasar dalam melakukan pemeriksaan dan penyelidikan kepada pelaku kejahatan.

Adanya surat dakwaan yang menguraikan penyebab dan faktor terjadinya tindak pidana dapat dimanfaatkan untuk memperoleh topik utama dari kumpulan dokumen surat dakwaan. Salah satu perkembangan teknologi yang berkaitan dengan *text* dapat dimanfaatkan untuk menemukan topik dari kumpulan kalimat menjadi fokus utama dalam penelitian ini. Metode yang dapat digunakan yaitu analisis *topic modelling*. Analisis *topic modelling* digunakan dalam menemukan tema utama yang meliputi kumpulan dokumen yang besar dan tidak terstruktur. Dalam melakukan analisis *topic modelling* peneliti menggunakan metode *Latent Dirichlet Allocation (LDA)* yang mempunyai tujuan untuk meringkas, melakukan pengelompokan, menghubungkan maupun memproses data dokumen yang sangat besar. Hasil yang didapatkan dari metode *LDA* berupa daftar topik yang diberi bobot untuk masing-masing dokumen (Blei et al., 2003). Dengan menggunakan

analisis ini maka dapat diketahui topik-topik apa saja yang sering muncul, sehingga dapat memudahkan dalam mengetahui dan memahami jenis kasus tindak pidana yang sering terjadi dan dibahas berdasarkan klasifikasi perkara di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.

## **1.2. Rumusan Masalah**

Berdasarkan latar belakang diatas maka rumusan masalah yang akan diangkat adalah sebagai berikut :

1. Bagaimana gambaran umum dari data kasus tindak pidana di Pengadilan Negeri Yogyakarta pada tahun 2019-2021?
2. Bagaimana implementasi *topic modelling* menggunakan *Latent Dirichlet Allocation (LDA)* berdasarkan surat dakwaan mengenai kasus tindak pidana yang sering terjadi dari 3 klasifikasi perkara tertinggi di Pengadilan Negeri Yogyakarta pada tahun 2019-2021?

## **1.3. Batasan Masalah**

Batasan masalah yang digunakan dalam penelitian ini agar masalah yang diangkat tidak menyimpang dari permasalahan penelitian sebagai berikut :

1. Pada penelitian ini data yang digunakan merupakan data sekunder dari Pengadilan Negeri Yogyakarta pada tahun 2019-2021 yang dikumpulkan pada tanggal 20 Desember 2022 – 10 Januari 2023.
2. Data yang digunakan merupakan surat dakwaan terkait kasus tindak pidana di Pengadilan Negeri Yogyakarta pada tahun 2019-2021 yang diambil dari laman <https://sipp.pn-yogyakarta.go.id/>.
3. Data akan diolah menggunakan *Software Microsoft Excel* dan *Phyton* menggunakan analisis *Latent Dirichlet Allocation (LDA)*.

## **1.4. Tujuan Penelitian**

Berdasarkan rumusan masalah diatas, maka tujuan dilaksanakannya penelitian ini adalah sebagai berikut :

1. Mengetahui gambaran umum dari data kasus tindak pidana di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.



2. Mengetahui hasil implementasi *topic modelling* menggunakan metode *Latent Dirichlet Allocation (LDA)* berdasarkan surat dakwaan mengenai kasus tindak pidana yang sering terjadi di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.

### **1.5. Manfaat Penelitian**

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut :

1. Menambah wawasan mengenai analisis *topic modelling* dengan menggunakan metode *Latent Dirichlet Allocation (LDA)* mengenai dokumen surat dakwaan di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.
2. Penelitian ini diharapkan dapat memberikan informasi mengenai jenis kasus tindak pidana yang sering terjadi di Kota Yogyakarta pada tahun 2019-2021 kepada aparat penegak hukum dan masyarakat umum.
3. Dapat dijadikan referensi bagi penelitian berikutnya terkait *text mining* dan *topic modelling* menggunakan *Latent Dirichlet Allocation* mengenai dokumen surat dakwaan.

## **BAB II**

### **TINJAUAN PUSTAKA**

Berbagai penelitian terdahulu dapat dijadikan referensi dan tolak ukur peneliti dalam menuliskan dan memperdalam teori yang akan digunakan dalam penelitian ini. Terdapat beberapa penelitian terdahulu yang berkaitan dengan penelitian yang dilakukan oleh peneliti, sebagai berikut.

Penelitian dilakukan oleh Sahria & Hatta Fudholi (2020) yang berjudul “Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode *Topic Modeling LDA (Latent Dirichlet Allocation)*”. Tujuan dari penelitian ini adalah mengetahui bagaimana metode *topic modelling LDA* melakukan analisis dari kumpulan judul penelitian Kesehatan dari jurnal SINTA. Hasil dari penelitian ini adalah terdapat 2 topik dominan yaitu topik umum dan topik penyakit, dengan hasil pengujian sebanyak 94,1% mengatakan sangat baik dan 5,1% mengatakan baik.

Penelitian dilakukan oleh Setiojohatmo et al (2020) yang berjudul “Analisis Metoda *Latent Dirichlet Allocation* untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik”. Tujuan dari penelitian ini adalah mendapatkan referensi dokumen Tugas Akhir untuk Jurusan Teknik Komputer dari kumpulan judul-judul Tugas Akhir. Hasil dari penelitian ini adalah mendapatkan referensi kumpulan dokumen Tugas Akhir yang terbagi menjadi 3 skenario yaitu skenario S1 dari 4 dokumen terdiri dari 2 topik, skenario S2 dari 60 dokumen terdiri dari 3 topik dan skenario S3 dari 80 dokumen terdiri dari 4 topik.

Penelitian dilakukan oleh Rakhmawati et al (2021) yang berjudul “Klasterisasi Topik Konten *Channel Youtube Gaming* Indonesia Menggunakan *Latent Dirichlet Allocation*”. Tujuan dari penelitian ini adalah mengidentifikasi topik-topik dari genre *gaming* yang ada di kanal *youtube gaming* Indonesia. Hasil dari penelitian ini terdapat 5 topik paling dominan yaitu bocoran *update* dan *season* baru, *review hero* dan *skin* baru, *live* bermain *game* saat ada *event* menarik, membahas *bug* yang terjadi pada *patch* baru rilis, dan konten memborong *skin*.

Penelitian dilakukan oleh Sastriana et al (2021) yang berjudul “*Implementation Of Topic Modelling With Latent Dirichlet Allocation (LDA) On The Judge’s Decision*”. Tujuan dari penelitian ini adalah mengidentifikasi topik-

topik dari putusan hakim terkait perkara pidana di Pengadilan Negeri Selong Kelas 1 B pada tahun 2015-2021. Hasil dari penelitian ini adalah mendapatkan 4 topik perkara pidana yaitu pencurian dengan kekerasan, pencurian dengan pemberatan, pencurian biasa, dan perjudian.

Penelitian dilakukan oleh Pratama et al (2022) yang berjudul “*Telkom University Opinion Topic Modeling On Twitter Using Latent Dirichlet Allocation During Covid-19 Pandemic*”. Tujuan dari penelitian ini adalah mengetahui pendapat masyarakat dari cuitan pengguna *twitter* terkait Universitas Telkom. Hasil dari penelitian ini adalah terdapat 9 topik dengan nilai alpha 0,31, nilai beta 0.01 dan nilai koherensi paling optimum -15,33029, serta topik yang paling banyak dibahas yaitu penerimaan mahasiswa baru.

Penelitian dilakukan oleh Kannitha et al (2022) yang berjudul “*Pemodelan Topik Pada Keluhan Pelanggan Menggunakan Algoritma Latent Dirichlet Allocation dalam Media Sosial Twitter*”. Tujuan dari penelitian ini adalah mengetahui keluhan pelanggan terkait penggunaan layanan internet pada perusahaan *First Media* dan *Indihome* di waktu Pembatasan Sosial Berskala Besar (PSBB), dalam upaya menekan penyebaran virus *Covid-19*. Hasil dari penelitian ini menunjukkan bahwa topik yang banyak dikeluhkan pelanggan yaitu internet ssering mati dan internet sering mati dan putus.

Penelitian dilakukan oleh Tresnasari et al (2020) yang berjudul “*Social Child Case Document Clustering based on Topic Modeling using Latent Dirichlet Allocation*”. Tujuan dari penelitian ini adalah mengelompokkan dokumen kasus sosial anak berdasarkan kesamaan deskripsi kasus dan menemukan topik sesuai dengan tipe kasus yang ditargetkan. Hasil dari penelitian ini menunjukkan bahwa terdapat 5 topik sesuai dengan tipe kasus yang ditargetkan dan penelitian ini mendukung proses penggunaan kembali pengetahuan tentang penanganan kasus sosial anak.

**Tabel 2.1** Deskripsi Penelitian Terdahulu

No	Nama Peneliti	Judul Penelitian	Metode	Persamaan	Perbedaan
1.	Yoga Sahria, Dhomas Hatta Fudholi (2020)	Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode <i>Topic Modeling LDA (Latent Dirichlet Allocation)</i>	<i>Latent Dirichlet Allocation (LDA)</i>	Menggunakan metode yang sama yaitu <i>Latent Dirichlet Allocation</i>	Data yang digunakan yaitu data kumpulan judul dari jurnal penelitian kesehatan di Indonesia
2.	Urip T. Setojohatmo, Setiadi Rachmat, Tati Susilawati, Yuda Rahman (2020)	Analisis Metoda <i>Latent Dirichlet Allocation</i> untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik	<i>Latent Dirichlet Allocation (LDA)</i>	Menggunakan metode yang sama yaitu <i>Latent Dirichlet Allocation (LDA)</i>	Data yang digunakan yaitu data kumpulan dokumen laporan Tugas Akhir
3.	Arief Rahman, Rekyan Bayu Waskitho, Muhammad Fajrul Alam Ulin Nuha, Nur Aini Rakhmawati (2021)	Klasterisasi Topik Konten <i>Channel Youtube Gaming</i> Indonesia Menggunakan <i>Latent Dirichlet Allocation</i>	<i>Latent Dirichlet Allocation (LDA)</i>	Menggunakan metode yang sama yaitu <i>Latent Dirichlet Allocation (LDA)</i>	Data yang digunakan yaitu data kumpulan video dari 10 kanal <i>gaming</i> terkenal di Indonesia
4.	Hanipar Mahyulis Sastriana, Wiwit Pura	<i>Implementation Of Topic Modelling With Latent Dirichlet</i>	<i>Latent Dirichlet</i>	Menggunakan metode yang sama	Data yang digunakan yaitu data putusan hakim

No	Nama Peneliti	Judul Penelitian	Metode	Persamaan	Perbedaan
	Nurmayanti, Lukmanul Hakim (2021)	<i>Allocation (LDA) On The Judge's Decision</i>	<i>Allocation (LDA)</i>	yaitu <i>Latent Dirichlet Allocation (LDA)</i>	di Pengadilan Negeri Selong Kelas 1 B
5.	Tandya Rizky Pratama,, Donni Richasdy, Mahendra Dwifebri Purbalakson (2022)	<i>Telkom University Opinion Topic Modeling on Twitter Using Latent Dirichlet Allocation During Covid-19 Pandemic</i>	<i>Latent Dirichlet Allocation (LDA)</i>	Menggunakan metode yang sama yaitu <i>Latent Dirichlet Allocation (LDA)</i>	Data yang digunakan diperoleh dari cuitan pengguna <i>twitter</i> pada masa pandemi <i>covid-19</i>
6.	Diandra Zakeshia Tiara Kannitha, Mustafid, Puspita Kartikasari (2022)	Pemodelan Topik Pada Keluhan Pelanggan Menggunakan Algoritma <i>Latent Dirichlet Allocation</i> Dalam Media Sosial <i>Twitter</i>	<i>Latent Dirichlet Allocation (LDA)</i>	Menggunakan metode yang sama yaitu <i>Latent Dirichlet Allocation</i>	Data yang digunakan diperoleh dari cuitan pengguna <i>twitter</i> terkait keluhan pada provider <i>Indihome</i> dan <i>First Media</i>
7.	Nur Annisa Tresnasari, Teguh Bharata Adji, Adhistya Erna Permanasari (2020)	<i>Social Child Case Document Clustering based on Topic Modeling using Latent Dirichlet Allocatiom</i>	<i>Latent Dirichlet Allocation (LDA)</i>	Menggunakan metode yang sama yaitu <i>Latent Dirichlet Allocation</i>	Data yang digunakan diperoleh dari kumpulan dokumen <i>SCC</i> Dinas Sosial Daerah Istimewa Yogyakarta

## **BAB III**

### **LANDASAN TEORI**

#### **3.1. Pidana**

Tindak pidana berasal dari terjemahan bahasa Belanda yang terdiri dari tiga kata yaitu "*straf*", "*baar*" dan "*feit*". *Straf* memiliki arti pidana dan hukum. *Baar* memiliki arti dapat atau boleh. *Feit* memiliki arti tindak, peristiwa, pelanggaran dan perbuatan. Sehingga tindak pidana merupakan perbuatan yang dilarang hukum dengan disertai ancaman dan sanksi berupa pidana jika melanggar (Chazawi, 2008). Pelaku tindak pidana akan mendapatkan sanksi pidana berupa pemidanaan. Sanksi pidana merupakan suatu ancaman atau akibat yang didapatkan dari perbuatan yang dilakukan. Tujuan diberikan sanksi pidana yaitu untuk mengurangi kasus kejahatan, pelanggaran dan ancaman yang dapat merugikan (Mertokusuma, 1986). Pemidanaan merupakan tindakan balas dendam kepada pelaku kejahatan yang melanggar hukum, mempunyai tujuan agar pelaku tidak mengulangi perbuatan tersebut (Gunarto, 2009). Menurut Sholehuddin (2003) pemidanaan mempunyai tujuan yaitu :

1. Pemidanaan memberikan kesadaran dan pengingat kepada pelaku kejahatan agar tidak mengulang perbuatan kejahatan yang sama.
2. Pemidanaan sebagai sarana penyembuhan kepada pelaku kejahatan agar mempunyai sikap yang baik dan dapat berbaur kembali dengan masyarakat secara wajar.
3. Pemidanaan sebagai tempat pendidikan moral agar pelaku kejahatan dapat menyadari dan mengakui kesalahan yang dilakukan.

##### **3.1.1 Jenis-Jenis Pidana**

Pidana di Indonesia yang diatur oleh KUHP dalam Pasal 10 terdiri dari pidana pokok dan pidana tambahan. Pidana pokok merupakan jenis pidana yang diperintahkan atau dijatuhkan langsung oleh hakim. Sedangkan, pidana tambahan merupakan jenis pidana yang di jatuhkan bersamaan dengan pidana pokok (Wahyuni, 2017). Berikut adalah jenis-jenis pidana berdasarkan pidana pokok dan pidana tambahan :

## A. Pidana Pokok

### a. Pidana Mati

Pidana mati dalam Pasal 10 KUHP adalah pidana terberat atas suatu hukuman dari kejahatan yang telah dilakukan pelaku. Para pelaku pada pidana mati akan dijatuhkan hukuman berupa pencabutan nyawa sesuai dengan putusan pengadilan (Wahyuni, 2017). Kejahatan yang dapat dijatuhkan pidana mati yaitu seperti membunuh kepala negara dan pembunuhan berencana.

### b. Pidana Penjara

Pidana penjara adalah pidana yang sering dijatuhkan atas suatu hukuman dengan memberikan batasan kepada pelaku kejahatan, dilakukan dengan mengurung pelaku di lembaga pemasyarakatan (Lamintang, 1984).

### c. Pidana Kurungan

Pidana kurungan adalah pidana yang bersifat ringan dibandingkan penjara. Pidana kurungan dijatuhkan atas suatu hukuman dengan memberikan batasan kepada pelaku kejahatan yang bersifat sementara (Lamintang, 1984).

### d. Pidana Denda

Pidana denda adalah pidana yang dijatuhkan atas suatu hukuman yang dilakukan oleh pelaku kejahatan dengan membayar sejumlah uang. Terdapat kategori pidana denda yang sering digunakan berdasarkan Pasal 79 Ayat 1 Undang-Undang No. 1 Tahun 2023 yaitu sebagai berikut.

1. Kategori 1 yaitu berkisar Rp 1.000.000
2. Kategori 2 yaitu berkisar Rp 10.000.000
3. Kategori 3 yaitu berkisar Rp 50.000.000
4. Kategori 4 yaitu berkisar Rp 200.000.000
5. Kategori 5 yaitu berkisar Rp 500.000.000
6. Kategori 6 yaitu berkisar Rp 2.000.000.000
7. Kategori 7 yaitu berkisar Rp 5.000.000.000
8. Kategori 8 yaitu berkisar Rp 50.000.000.000

## B. Pidana Tambahan

### a. Pencabutan Hak Tertentu

Pencabutan hak tertentu adalah pidana tambahan yang dijatuhkan atas suatu hukuman yang dilakukan oleh pelaku kejahatan dengan tidak diperkenankan seperti menjalankan suatu jabatan atau kekuasaan.

### b. Perampasan Barang Tertentu

Perampasan barang tertentu adalah pidana tambahan yang dijatuhkan atas suatu hukuman yang dilakukan oleh pelaku kejahatan dengan merampas barang-barang yang dihasilkan dari kejahatan.

### c. Pengumuman Putusan Hakim

Pengumuman putusan hakim adalah pidana tambahan berupa pengumuman putusan yang telah dilakukan pelaku kejahatan kepada masyarakat, biasanya putusan akan dipublikasikan dalam bentuk koran dan media elektronik.

## 3.1.2 Jenis-Jenis Acara Pemeriksaan Pidana

Jenis acara pemeriksaan pidana di Indonesia berdasarkan Undang-Undang No. 8 Tahun 1981 tentang Hukum Acara Pidana (Indroharto, 2003) terdiri dari 3 yaitu, sebagai berikut :

### A. Acara Pemeriksaan Biasa

Acara pemeriksaan biasa merupakan pemeriksaan yang biasa dilakukan terhadap perkara yang membutuhkan pembuktian dan penerapan hukum yang tidak mudah dan sederhana. Tujuan dari pemeriksaan ini adalah mendapatkan suatu putusan pengadilan yang final, baik dan berkualitas.

### B. Acara Pemeriksaan Cepat

Acara pemeriksaan cepat merupakan pemeriksaan yang dilakukan dengan penyingkatan batas waktu atau penyederhanaan unsur-unsur yang terdapat acara pemeriksaan biasa. Tujuan dari pemeriksaan ini adalah mendapatkan suatu putusan yang lebih cepat dari sengketa yang diajukan.

### C. Acara Pemeriksaan Singkat

Acara pemeriksaan singkat merupakan pemeriksaan yang dilakukan dengan pemutusan suatu sengketa secara singkat. Tujuan dari pemeriksaan ini adalah mendapatkan suatu putusan yang sifatnya sederhana.



### **3.2. Surat Dakwaan**

Surat dakwaan menurut Abdul Karim Nasution merupakan surat atau akta yang berisi rumusan dari tindak pidana yang didakwakan kepada terdakwa, yang hasilnya dapat diambil sebagai kesimpulan sementara dari pemeriksaan yang akan menjadi dasar bagi hakim melakukan pemeriksaan jika bukti tercukupi sehingga terdakwa dapat dijatuhi hukuman (Nasution, 1972). Surat dakwaan mempunyai peranan yang sangat penting dalam perkara pidana diantaranya yaitu sebagai dasar dalam pemeriksaan dan penyelidikan, dasar dalam memberikan tuntutan pidana, dasar dalam melakukan pembelaan dan dasar bagi hakim dalam menjatuhkan putusan (Marpaung, 1992).

Dilihat dari berbagai kepentingan berdasarkan Surat Dakwaan Jaksa Agung Republik Indonesia No : SE-004/JA/11/1993 mempunyai berbagai fungsi yang berkaitan dengan pemeriksaan perkara pidana, sebagai berikut :

- a. Bagi pengadilan, surat dakwaan dijadikan dasar dalam pemeriksaan dan pertimbangan dalam pengambilan keputusan.
- b. Bagi penuntut umum, surat dakwaan dijadikan dasar pembuktian tuntutan pidana dan penggunaan upaya hukum.
- c. Bagi terdakwa, surat dakwaan dijadikan dasar dalam mempersiapkan pembelaan.

### **3.3. Statistika Deskriptif**

Statistika deskriptif merupakan metode yang berkaitan dengan pengumpulan atau penyajian data yang dapat memberikan informasi berguna. Terdapat dua jenis statistik yaitu statistik deskriptif dan statistik inferensial. Dengan menggunakan statistika deskriptif berbagai kumpulan data data ditampilkan secara ringkas dan jelas, serta informasi dasar dari kumpulan data yang ada dapat disediakan. Informasi yang diperoleh dari statistik deskriptif antara lain terdiri dari ukuran kumpulan data, ukuran sebaran data, dan tren kumpulan data (Mutsani, 2019).

### **3.4. Machine Learning**

*Machine learning* atau pembelajaran mesin merupakan suatu pendekatan dalam *Artificial Intelligence* yang banyak digunakan untuk menggantikan atau meniru perilaku manusia dalam menyelesaikan masalah (Ahmad, 2017). Pada prosesnya, *machine learning* dapat melakukan prediksi data dalam jumlah besar atau data sampel yang akan dimodifikasi dan presentasikan dalam sebuah keputusan.

*Machine learning* mempunyai empat proses algoritma (Sinaga, 2020), sebagai berikut :

a. *Supervised Learning*

Algoritma yang digunakan untuk mengidentifikasi hubungan antara input data dan label *output*.

b. *Unsupervised Learning*

Algoritma yang digunakan untuk mengumpulkan informasi terkait data yang tidak berlabel.

c. *Semi-Supervised Learning*

Algoritma ini merupakan gabungan dari *supervised learning* dan *unsupervised learning*, digunakan untuk mengidentifikasi data yang berlabel dan tidak berlabel.

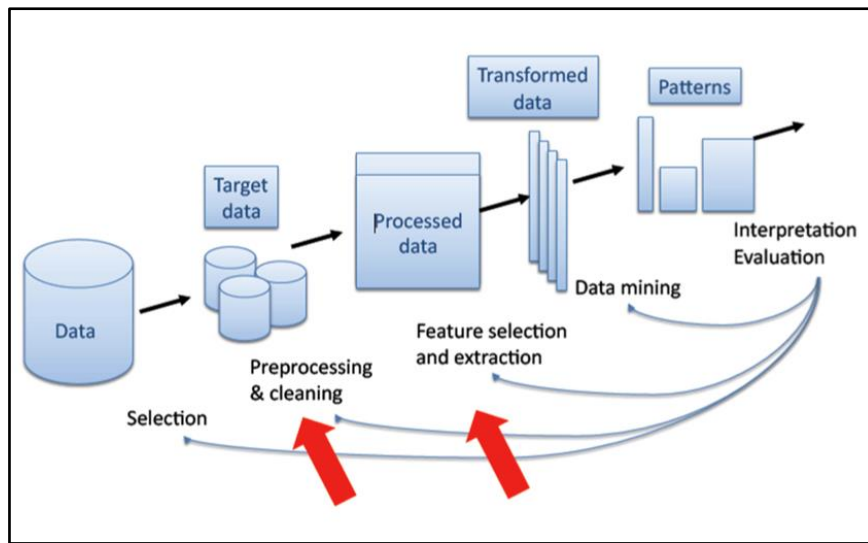
d. *Reinforcement Learning*

Algoritma yang digunakan dalam menentukan sebuah keputusan yang tepat untuk mendapatkan kinerja algoritma yang maksimal.

### **3.5. Text Mining**

*Text Mining* adalah suatu proses yang dilakukan untuk mendapatkan informasi berkualitas tinggi dari teks dengan melihat pola dan tren (Agatha Deolika, 2019 ). *Text Mining* berkaitan dengan pengembangan berbagai ilmu matematika, statistik, linguistik dan pengenalan pola yang memungkinkan proses analisis secara otomatis dari sebuah data yang tidak terstruktur (Scholtes, 2009). Data yang tidak terstruktur adalah data yang dapat dikatakan bersifat kualitatif dan tidak memiliki format tertentu, seperti foto, gambar, video dan audio. Sedangkan data terstruktur adalah data yang memiliki format tertentu dan memiliki kemiripan satu sama lain, seperti hasil survei dan kuisioner. *Text Mining* memiliki lima tahapan yaitu *selection*,

*preprocessing and cleaning, transformation, data mining dan interpretation and evaluation.* Tahapan *text mining* digambarkan sebagai berikut (Gata, 2017).



**Gambar 3.1** Proses *Text Mining*

### 3.6. *Text Processing*

Peran yang sangat penting dalam proses *text mining* terdapat pada *preprocessing*. *Preprocessing* berperan penting dalam menghilangkan sebagian teks yang tidak diperlukan guna mendapatkan data berkualitas dan akurat yang memenuhi syarat pada algoritma (Suwanto Sanjaya, 2015). *Preprocessing* mempunyai beberapa tahapan sebagai berikut :

a. *Case Folding*

*Case folding* adalah kegiatan menyamakan atau mengkonversi keluruhan teks dalam dokumen dengan mengubah huruf kapital menjadi huruf kecil (Debby Alita, 2020). Pada proses ini beberapa karakter yang tidak berhubungan dengan huruf akan dihilangkan, karena hanya menerima karakter “a sampai “z”.

b. *Filtering/ Removing Punctuation and Numbers*

*Filtering/ Removing Punctuation and Numbers* adalah kegiatan membuang atau membersihkan data yang dianggap tidak penting seperti tanda baca, angka dan special karakter lainnya seperti “!, \$, %, &, @, \*, #, {, }” atau biasanya dikenal dengan *noise*. *Noise* merupakan suatu bentuk data yang nantinya dapat mengganggu proses pengolahan data (Cendana & Permana, 2019).

c. *Stopword*

*Stopword* adalah kegiatan menghapus kata yang sering muncul dalam dokumen namun dianggap tidak memiliki makna, sehingga dapat dihilangkan. Pada proses ini beberapa kata yang sering muncul seperti “dan”, “di”, “ke”, “yang”, “dengan”, “atau” dan sebagainya (Socrates et al., 2016).

d. *Tokenizing*

*Tokenizing* adalah kegiatan memisahkan kalimat ke dalam unit yang lebih kecil serta berfungsi dalam membersihkan beberapa karakter tertentu yang dianggap sebagai tanda baca (Hikmah, 2018). Pada proses ini dokumen dipisahkan menjadi potongan-potongan token yang terdiri dari angka, simbol, tanda baca dan entitas penting lainnya. *Tokenizing* sangat diperlukan agar dalam tahapan selanjutnya perhitungan menjadi lebih mudah.

e. *Stemming*

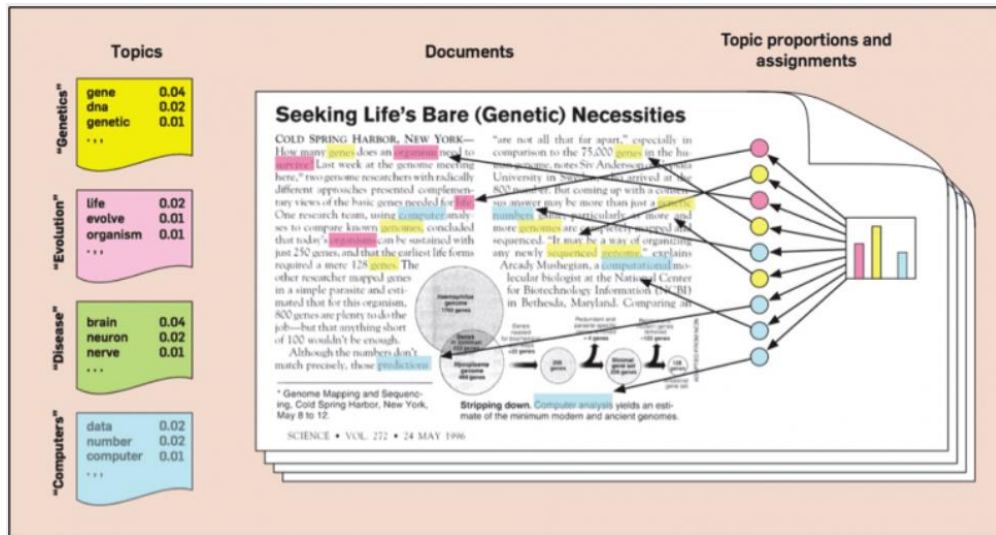
*Stemming* adalah kegiatan mengubah sebuah kata menjadi kata dasar dengan menghapus awal dan akhiran kata yang mempunyai imbuhan (Hikmah, 2018). Hasil dari *stemming* akan berdampak pada proses *text mining* selanjutnya, karena dapat mempengaruhi baik tidaknya pada perhitungan suatu teks.

### 3.7. *Topic Modelling*

*Topic Modelling* atau pemodelan topik merupakan metode *unsupervised learning* yang menerapkan pengelompokan dari kumpulan teks besar untuk menemukan sebuah tema (Nurlayli, 2019). Menurut Blei (2003) *Topic Modeling* mempunyai konsep yang terdiri dari “kata”, “dokumen” dan “corpora”. Konsep “kata” diartikan sebagai item dasar pada dokumen, “dokumen” diartikan sebagai banyaknya susunan kata, dan “corpora” diartikan sebagai kumpulan teks yang didapatkan dari susunan kata yang terdapat pada dokumen (Blei et al., 2003). Algoritma *Topic Modelling* dalam menganalisis kata-kata pada suatu teks akan dikumpulkan menjadi satu, yang digunakan untuk menemukan suatu topik atau tema.

*Topic Modeling* banyak digunakan dalam proses memisahkan data untuk menemukan topik tersembunyi dari suatu dokumen yang mempunyai proporsi

tertentu sesuai dengan kata-kata yang tercantum didalamnya. Hasil dari *Topic Modeling* dapat digunakan dalam membuat rangkuman, membuat visualisasi dan melakukan eksplorasi dan berteori tentang *corpus*. Konsep *Topic Modeling* menurut Blei (2010) dapat dilihat pada gambar berikut.



**Gambar 3.2** Konsep *Topic Modelling*

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

*Term Frequency – Inverse Document Frequency (TF-IDF)* merupakan metode pembobotan antara suatu kata (*term*) terhadap dokumen yang bertujuan untuk mengubah data berbentuk teks menjadi numerik (Harjanta, 2015). *TF-IDF* merupakan gabungan dari dua metode yaitu *Term Frequency (TF)* dan *Inverse Document Frequency (IDF)*. Metode *TF* bertujuan untuk menghitung nilai dari setiap kemunculan *term(t)* dalam sebuah dokumen. Semakin sering *term(t)* muncul dalam sebuah dokumen, maka semakin besar nilai *TF*. Metode *IDF* bertujuan untuk menghitung nilai dari setiap kemunculan *term(t)* pada semua dokumen. Semakin sering *term(t)* muncul dalam beberapa dokumen, maka semakin umum *term(t)* tersebut sehingga tidak boleh digunakan (Nugroho et al., 2021).

*Term Frequency (TF)* memiliki beberapa jenis algoritma sebagai berikut (Yoren, 2018) :

1. *Binary Term Frequency (TF Binary)*, yaitu penyeragaman bobot pada dokumen yang terdiri dari 0 jika *term* tidak muncul dan 1 jika *term* muncul
2. *Raw Term Frequency (TF Murni)*, yaitu nilai *TF* yang didapatkan berdasarkan frekuensi dari suatu *term* pada dokumen. jika muncul *term* sebanyak 3 kali, maka *term* tersebut akan bernilai 3.

3. *Logarithmic Term Frequency (TF Logaritmik)*, yaitu nilai *TF* dengan fungsi logaritmik untuk perhitungannya menghindari dokumen yang mengandung sedikit *term*, namun memiliki frekuensi yang tinggi. Berikut merupakan rumus *TF* logaritmik.

$$TF_{dt} = 1 + \log (f_{dt}) \quad (3.2)$$

Keterangan :

$f_{dt}$  = frekuensi kemunculan kata ( $t$ ) dalam dalam dokumen ( $d$ )

4. *Augmented Term Frequency (TF Normalisasi)*, yaitu perhitungan nilai *TF* menggunakan perbandingan antara frekuensi sebuah kata dengan jumlah keseluruhan kata pada dokumen. Berikut merupakan rumus *TF* normalisasi.

$$TF_{dt} = 0.5 + 0.5 \left( \frac{TF}{\max Tf} \right) \quad (3.3)$$

Dimana nilai  $\max TF$  merupakan jumlah muncul *term(t)* terbanyak pada dokumen yang sama.

Perhitungan *IDF* digunakan untuk mengetahui hubungan *term(t)* dalam sebuah dokumen, dengan nilai *IDF* akan semakin besar jika jumlah dokumen yang mengandung *term(t)* semakin sedikit. Berikut merupakan rumus dari *IDF*.

$$IDF_{dt} = \ln \left( \frac{D}{DF_t} \right) + 1 \quad (3.4)$$

Keterangan :

$IDF_t$  = *Inverse Document Frequency*

$D$  = Total semua dokumen

$DF_t$  = Frekuensi dokumen dari *term (t)*

Perhitungan *TF* digunakan untuk mengetahui bobot pada setiap *term* dalam sebuah dokumen. Berikut merupakan rumus dari *TF* yaitu  $TF(W_{dt})$ .

$$W_{dt} = TF_{dt} \times IDF_t \quad (3.5)$$

$$W_{dt} = TF_{dt} \times \left( \ln \left( \frac{D}{DF_t} \right) \right) \quad (3.6)$$

Keterangan :

$W_{dt}$  = Bobot *term (t)* terhadap dokumen

$TF_{dt}$  = Frekuensi kemunculan *term (t)* dalam dokumen

$D$  = Total dokumen

$IDF_{dt}$  = *Inverse document frequency*

Berapapun nilai  $TF_{dt}$  apabila  $D = DF_t$  maka hasilnya akan bernilai 0. Hal itu dikarenakan nilai dari  $\ln(1) = 0$  untuk perhitungan  $IDF$ . Sehingga besarnya nilai  $TF_{dt}$  tidak akan memberi pengaruh dalam perhitungan bobot. Oleh karena itu, perlu menambahkan nilai 1 pada perhitungan bobotnya, seperti rumus berikut.

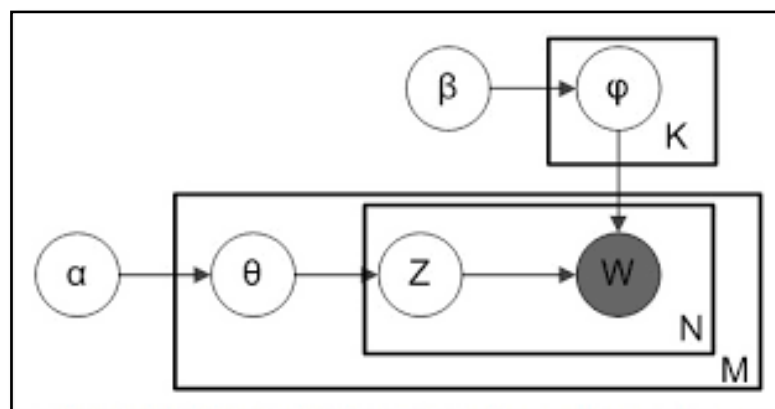
$$W_{dt} = TF_{dt} \times \left( \ln \left( \frac{D}{DF_t} \right) + 1 \right) \quad (3.7)$$

### 3.9. Latent Dirichlet Allocation (LDA)

*Latent Dirichlet Allocation (LDA)* merupakan model *probabilistic generative* dari sebuah *corpus*. *LDA* mempunyai dasar bahwa sebuah dokumen direpresentasikan sebagai campuran acak topik laten, dengan setiap topik ditandai dengan distribusi atas kata-kata. *LDA* adalah salah satu algoritma *Topic Modelling* yang banyak digunakan untuk melakukan analisis pada teks dengan jumlah yang besar (Blei et al., 2003). *LDA* memberikan representasi makna dari sebuah dokumen dalam *corpus* dengan mengasumsikan proses generatif sebagai berikut :

1. Topik dipilih secara acak berdasarkan distribusi dari setiap dokumen.
2. Kata dipilih dari distribusi yang berhubungan dengan topik yang dipilih.

Metode *LDA* dalam pemodelan topiknya dilakukan tanpa pengawasan dan tidak memerlukan ringkasan yang mendetail terkait suatu data (metadata) untuk pembelajaran, sehingga dokumen yang mempunyai jumlah besar dapat dianalisis dalam waktu singkat. Hal itu menjadikan metode *LDA* sering digunakan dalam analisis pemodelan topik. Representasi model grafis *LDA* menurut Setiojohatmo (2020) seperti berikut.



**Gambar 3.3** Representasi Model Grafis *LDA*

Pada **Gambar 3.3** merupakan representasi model grafis *LDA* yang mempunyai 3 tingkatan. Tingkatan pertama terdapat pada  $\alpha$  dan  $\beta$  merupakan parameter distribusi topik yang diasumsikan dengan penggunaan sampel satu kali untuk menghasilkan *corpus*. Parameter  $\alpha$  digunakan untuk menentukan distribusi topik dalam dokumen dan  $\beta$  digunakan untuk menentukan distribusi kata dalam topik. Semakin tinggi nilai yang dihasilkan  $\alpha$  menunjukkan bahwa semakin banyak topik yang dibahas dalam suatu dokumen, dan semakin tinggi nilai yang dihasilkan  $\beta$  menunjukkan bahwa semakin banyak kata-kata yang terdapat dalam suatu topik. Pada variabel  $\theta$  merupakan variabel yang terdapat pada tingkatan dokumen (M), direpresentasikan sebagai distribusi topik untuk dokumen tertentu. Semakin tinggi nilai  $\theta$  menunjukkan semakin banyak topik yang terdapat dalam suatu dokumen. Pada variabel  $Z$  dan  $W$  merupakan variabel yang terdapat pada tingkatan kata (N). variabel  $Z$  direpresentasikan sebagai topik dari kata tertentu dalam suatu dokumen dan  $W$  direpresentasikan sebagai kata yang berkaitan dengan topik dalam suatu dokumen (Blei et al., 2003). Sehingga berdasarkan representasi model *LDA*, nilai probabilitas *LDA* dirumuskan sebagai berikut :

$$p(W, Z, \theta, \varphi | \alpha, \beta) = p(\varphi | \beta) p(\theta | \alpha) p(Z | \theta) p(W | \varphi_K) \quad (3.8)$$

Keterangan :

$p(W, Z, \theta, \varphi | \alpha, \beta)$  = Nilai Probabilitas *LDA*

$\alpha$  = *Dirichlet* parameter atas distribusi topik terhadap dokumen

$\beta$  = *Dirichlet* parameter atas distribusi kata terhadap topik

$\theta$  = Distribusi topik untuk dokumen tertentu

$\varphi$  = Distribusi kata untuk topik tertentu

$Z$  = Topik

$W$  = Kata

$K$  = Kumpulan topik

$N$  = Kumpulan kata

$M$  = Kumpulan dokumen

Metode analisis *Latent Dirichlet Allocation (LDA)* mempunyai tiga term yaitu *word*, *document* dan *corpus*. *Word* merupakan unit dasar, *document* merupakan susunan *word* dan *corpus* merupakan kumpulan *document*. Perhitungan *corpus* pada metode analisis *LDA* sulit dihitung secara manual karena mempunyai distribusi yang



sangat kompleks. Sehingga perhitungan dilakukan menggunakan algoritma *Gibbs Sampling* yang berfungsi dalam mengestimasi parameter model *LDA*. Berikut adalah rumus dari algoritma *Gibbs Sampling* (Setijohatmo et al., 2020).

$$\theta_{dw,k} = \frac{n_{-w,k}^w + \beta}{n_{-w,k} + W\beta} \frac{n_{-w,k}^{dw} + \alpha}{n_{-w}^{dw} + K\alpha} \quad (3.9)$$

Keterangan :

$\theta_{dw,k} = p(W, Z, \theta, \varphi | \alpha, \beta)$  = Nilai Probabilitas *LDA*

$\alpha$  = *Dirichlet* parameter atas distribusi topik terhadap dokumen

$\beta$  = *Dirichlet* parameter atas distribusi kata terhadap topik

$n_{-w,k}^w$  = Berapa kali kata  $w$  di *assign* ke topik  $k$  di setiap dokumen

$n_{-w,k}$  = Berapa kali kata selain  $w$  di *assign* ketopik  $k$  di setiap dokumen

$n_{-w,k}^{dw}$  = Berapa kali topik  $k$  di *assign* kedokumen  $d$

$n_{-w}^{dw}$  = Berapa kali topik selain  $k$  di *assign* ke  $d$

### 3.10. Topic Coherence

*Topic Coherence* atau koherensi topik adalah metode yang digunakan untuk mengukur nilai dari suatu topik berdasarkan tingkat kesamaan antara kata-kata yang terdapat didalam topik. Tujuan dari *Topic Coherence* yaitu menginterpretasikan topik-topik yang memiliki keterkaitan atau hubungan secara statistik (Putra & Kusumawardani, 2017).

*Topic Coherence* dinyatakan dalam sebuah nilai yaitu *Coherence Score*. *Coherence Score* yaitu ukuran yang digunakan untuk evaluasi dan menentukan banyaknya model dari *Topic Modelling* yang akan divisualisasi atau ditampilkan dalam bentuk grafik. Semakin besar *Coherence Score*, maka akan semakin baik model yang dihasilkan. *Coherence Score* dihitung menggunakan skor berpasangan (UCI atau UMass) pada setiap kata yang dipilih dan digabungkan dengan semua skor berpasangan untuk topik tertentu. *Coherence Score* dapat diperoleh dengan perhitungan UMass (Tresnasari et al., 2020) sebagai berikut :

$$score(v_i, v_j, \epsilon) = \log \frac{D(v_i, v_j) + \epsilon}{D(v_j)} \quad (3.10)$$

Keterangan :

$v$  = Kumpulan kata yang mengandung topik

$D(v_i, v_j)$  = Jumlah dokumen yang mengandung kata  $v_i$  dan  $v_j$

$D(v_j)$  = Jumlah dokumen yang mengandung kata  $v_j$

$\epsilon$  = Faktor pemulusan yang menjamin skor adalah bilangan asli

## BAB IV

### METODOLOGI PENELITIAN

#### 4.1. Populasi Penelitian

Populasi yang akan digunakan dalam penelitian ini adalah seluruh data kasus tindak pidana yang terdapat di Pengadilan Negeri Yogyakarta. Sampel penelitian ini adalah surat dakwaan kasus tindak pidana di Pengadilan Negeri Yogyakarta pada tahun 2019-2021 dengan total 971 kasus.

#### 4.2. Jenis Data dan Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder, dikumpulkan pada tanggal 20 Desember 2022 – 10 Januari 2023. Data bersumber dari Sistem Informasi Penelusuran Perkara Pengadilan Negeri Yogyakarta yaitu <https://sipp.pn-yogyakarta.go.id/>, dengan status perkara bersifat minutas (Sudah diputuskan) dan acara pemeriksaan pidana bersifat biasa.

#### 4.3. Variabel penelitian

Variabel yang digunakan dan definisi dari masing-masing variabel dalam penelitian ini dijelaskan pada **Tabel 4.1**, sebagai berikut.

**Tabel 4.1** Variabel Penelitian

No	Variabel Penelitian	Definisi Variabel Penelitian
1	Surat Dakwaan	Surat yang digunakan sebagai dasar dalam tahapan tuntutan pidana
2	Klasifikasi Perkara Pidana	Jenis perkara pidana yang dikelompokkan berdasarkan karakteristik tertentu
3	Denda	Sanksi pidana berupa sejumlah uang yang harus dibayar dalam jumlah tertentu
3	Hukuman	Sanksi pidana berupa penahanan seseorang dalam waktu tertentu

#### **4.4. Metode Analisis Data**

Penelitian ini dilakukan menggunakan *software Microsoft Excel, RStudio* dan *Phyton*. Adapun beberapa metode yang digunakan pada penelitian ini sebagai berikut.

1. Analisis Deskriptif

Analisis deskriptif digunakan untuk mengetahui gambaran umum dari data kasus tindak pidana di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.

2. *Text Mining*

*Text Mining* digunakan untuk memproses data yang tidak terstruktur menjadi lebih terstruktur yang dapat mewakili semua dokumen.

3. *TF-IDF*

*Term Frequency-Inverse Document Frequency (TF-IDF)* digunakan untuk melakukan proses pembobotan kata dari data teks menjadi data numerik.

4. *Topic coherence*

*Topic coherence* digunakan untuk menentukan jumlah topik *modelling* dengan melihat visualisasi pada grafik *coherence score*

5. *Topic Modelling* dengan *Latent Dirichlet Allocation (LDA)*

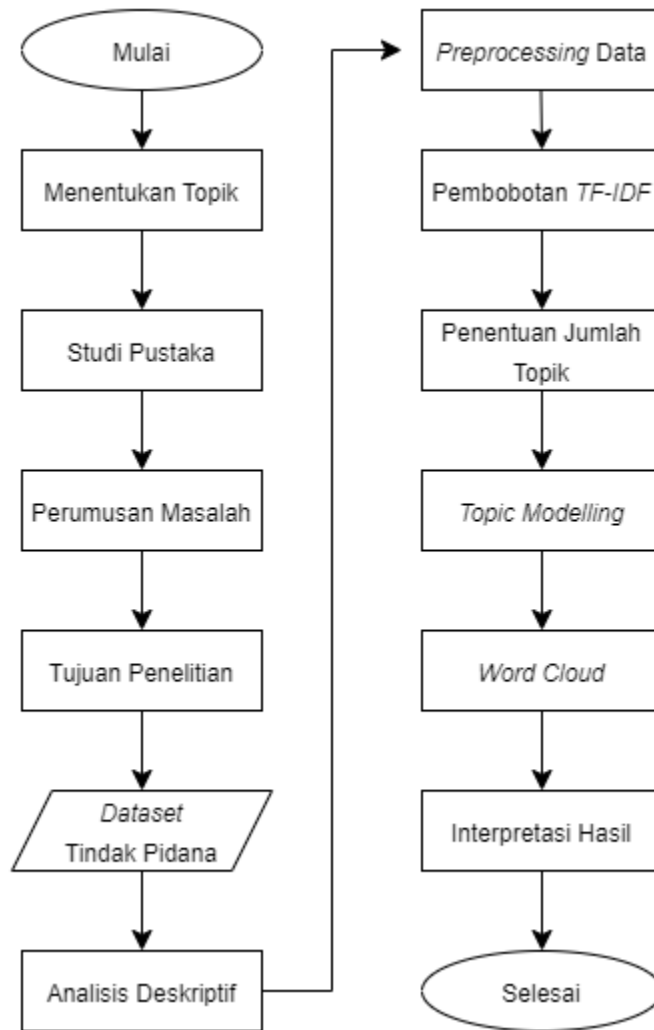
*Latent Dirichlet Allocation (LDA)* digunakan untuk mengelompokkan setiap kata menjadi topik dari setiap dokumen dan mencari hubungan antar kata.

6. *Wordcloud*

*Wordcloud* digunakan untuk menampilkan gambaran visual dari frekuensi kata pada teks atau dokumen.

#### **4.5. Tahapan Penelitian**

Tahapan penelitian pada penelitian ini dijelaskan pada gambar *flowchart* dibawah ini.



**Gambar 4.1** Diagram Alir Penelitian

Adapun tahapan penelitian berdasarkan **Gambar 4.1** adalah sebagai berikut :

1. Menentukan topik yang akan diangkat dalam penelitian.
2. Melakukan studi pustaka terkait topik penelitian untuk mencari informasi yang relevan dengan masalah yang akan diteliti.
3. Mengidentifikasi rumusan masalah dari penelitian.
4. Mengidentifikasi tujuan dari penelitian.
5. Mendapatkan *dataset* terkait kasus tindak pidana di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.
6. Melakukan *preprocessing* data sebagai tahap awal pemrosesan data yang terdiri dari tahapan *case folding*, *remove punctuation* dan *numbers*, *stopword*, *stemming* dan *tokenizing*.
7. Melakukan analisis deskriptif untuk melihat gambaran data kasus tindak pidana secara umum di Pengadilan Negeri Yogyakarta tahun 2019-2021.

8. Melakukan pembobotan *TF-IDF* untuk memberikan bobot hubungan pada suatu kata terhadap suatu dokumen.
9. Menentukan jumlah topik yang akan digunakan berdasarkan nilai *coherence* tertinggi.
10. Mendapatkan hasil pemodelan topik menggunakan metode *Latent Dirichlet Allocation (LDA)*.
11. Membuat visualisasi berupa *word cloud* untuk melihat gambaran teks yang sering muncul pada dokumen.
12. Melakukan interpretasi pemodelan topik yang dihasilkan.

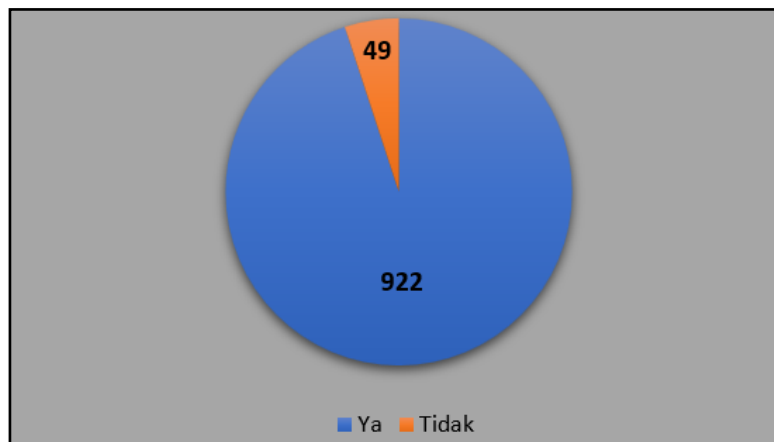
## BAB V

### HASIL DAN PEMBAHASAN

#### 5.1. Analisis Deskriptif dan Uji Independensi

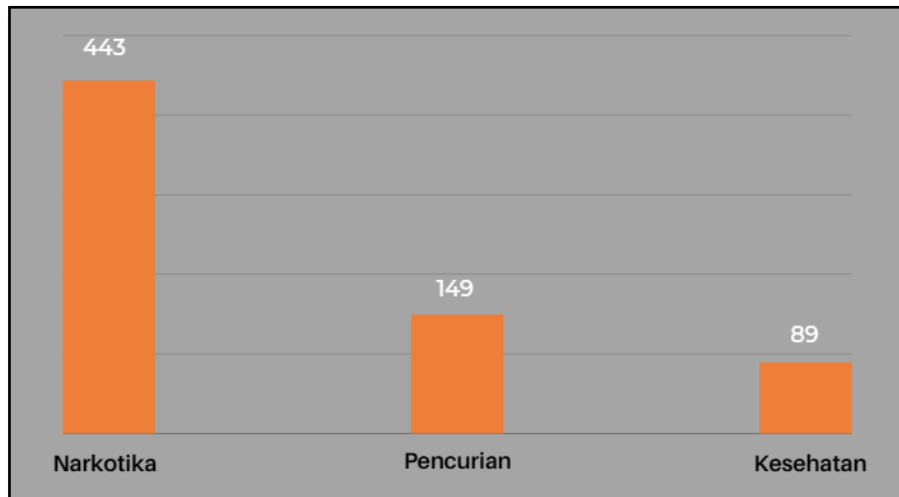
##### 5.1.1 Analisis Deskriptif

Analisis deskriptif dalam penulisan ini bertujuan untuk melihat gambaran secara umum dari data yang digunakan, yaitu data mengenai kasus tindak pidana di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.



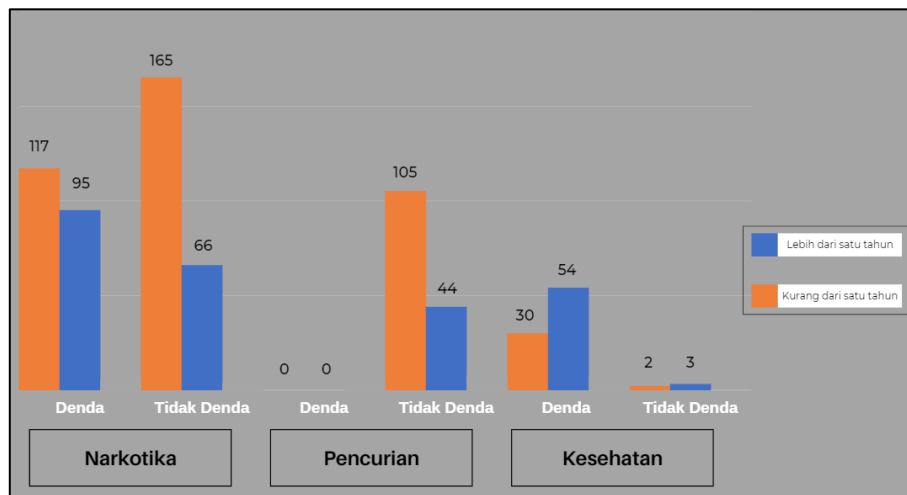
**Gambar 5.1** Diagram Publikasi Kasus Tindak Pidana

Pada **Gambar 5.1** merupakan grafik berupa diagram lingkaran mengenai publikasi kasus tindak pidana di Pengadilan Negeri Yogyakarta pada tahun 2019-2021. Terlihat bahwa dari total sebanyak 971 kasus tindak pidana terdapat 922 kasus tindak pidana terpublikasi atau dapat diakses oleh masyarakat luas, sedangkan 49 kasus tindak pidana tidak terpublikasi atau tidak dapat diakses oleh masyarakat luas. Berdasarkan data yang diperoleh di Pengadilan Negeri Yogyakarta tercatat bahwa kasus tindak pidana pada tahun 2019 terdapat sebanyak 325 kasus, pada tahun 2020 terdapat sebanyak 320 kasus dan pada tahun 2021 terdapat sebanyak 326 kasus.



**Gambar 5.2** Diagram Klasifikasi Perkara Kasus Tindak Pidana

Pada **Gambar 5.2** merupakan grafik berupa diagram batang mengenai klasifikasi perkara dari kasus tindak pidana yang sering terjadi di Pengadilan Negeri Yogyakarta pada tahun 2019-2021. Dari grafik dapat terlihat bahwa kasus tindak pidana dengan mengambil 3 klasifikasi perkara tertinggi terdiri dari narkotika dengan total kasus sebanyak 443, pencurian dengan total kasus sebanyak 149 dan kesehatan dengan total kasus sebanyak 89.



**Gambar 5.3** Diagram Denda dan Hukuman Kasus Tindak Pidana

Pada **Gambar 5.3** merupakan grafik berupa diagram batang mengenai hukuman dan denda kasus tindak pidana berdasarkan 3 klasifikasi tertinggi terdiri dari narkotika, pencurian dan kesehatan di Pengadilan Negeri Yogyakarta pada tahun 2019-2021. Dapat terlihat bahwa kasus tindak pidana tertinggi terdapat pada hukuman penjara waktu kurang dari satu tahun dengan tidak mendapatkan denda di klasifikasi perkara narkotika yaitu sebanyak 450 kasus. Pada kasus tindak pidana



hukuman dan denda saling berkaitan dikarenakan dalam pemberiannya harus mempertimbangkan berbagai faktor seperti kondisi ekonomi, berat dan ringannya kasus, serta kerugian yang ditimbulkan. Pemberian denda mempertimbangkan kesanggupan pelaku dalam membayar, jika pelaku tidak dapat membayar akan digantikan menjadi hukuman penjara sesuai durasi waktu yang ditentukan. Dalam kasus tindak pidana, hukuman penjara waktu banyak diberikan karena dapat menjamin keamanan dan menghindari timbulnya kerugian yang disebabkan pelaku.

Berdasarkan 3 klasifikasi perkara tertinggi tersebut akan dijadikan acuan dalam melakukan analisis *topic modelling* menggunakan dokumen surat dakwaan untuk mengetahui jenis kasus yang sering terjadi dan dibahas di Pengadilan Negeri Yogyakarta pada tahun 2019-2021. Tahapan selanjutnya yaitu melakukan *preprocessing* data menggunakan dokumen surat dakwaan.

## 5.2. *Preprocessing*

*Preprocessing* merupakan tahapan awal yang mempunyai tujuan untuk membersihkan atau menghilangkan teks pada data yang tidak diperlukan. *Preprocessing* terdiri dari beberapa tahapan diantaranya yaitu *case folding*, *remove punctuation* dan *numbers*, *stopword*, *stemming* dan *tokenizing*. Berikut contoh data dari dokumen surat dakwaan kasus tindak pidana yang digunakan dalam penelitian ini, sedangkan untuk data lengkap dapat dilihat dalam Lampiran.

**Tabel 5.1** Data Surat Dakwaan Pidana

Surat Dakwaan
Bahwa mereka terdakwa 1 BURHAN bin ABDUL SAMAD (alm) dan terdakwa 2 SADAKA bin BASRI BABBAI (alm) pada hari Senin tanggal 07 Juni 2021 sekitar jam 03.00 Wib atau setidak-tidaknya pada waktu lain dalam bulan Juni tahun 2021 bertempat di Kantor Yayasan “BIAS” Ruko Jl. Tri tunggal Kav 2 Sorosutan Kec Umbulharjo Kota Yogyakarta, atau setidak-tidaknya pada suatu tempat yang masih termasuk dalam daerah hukum Pengadilan Negeri Yogyakarta, para terdakwa telah mengambil barang sesuatu yang seluruhnya atau sebagian kepunyaan orang lain, dengan maksud untuk memiliki secara melawan hukum, pada waktu malam dalam sebuah rumah atau pekarangan tertutup yang ada rumahnya, yang dilakukan oleh orang yang adanya disitu tidak diketahui atau tidak dikehendaki oleh orang yang berhak, yang dilakukan oleh dua orang atau

lebih bersama-sama yang untuk masuk ketempat melakukan kejahatan, atau untuk sampai pada barang yang diambilnya dilakukan dengan membongkar, memecah, atau memanjat atau dengan memakai anak kunci palsu, perintah palsu atau pakaian jabatan palsu. Perbuatan para terdakwa sebagaimana diatur dan diancam pidana dalam Pasal 363 ayat (1) ke, 3, 4, dan 5 KUHP.

Terdapat beberapa tahapan dalam melakukan *preprocessing* data pada penelitian ini, yaitu terdiri dari proses *case folding*, *remove punctuation and numbers*, *stopword*, *stemming* dan *tokenizing*. Berikut adalah tahapan *preprocessing* data.

### 5.2.1 Case Folding

*Case folding* merupakan tahapan yang bertujuan untuk mengubah semua huruf besar/kapital (*uppercase*) menjadi huruf kecil (*lowercase*) dari huruf “a” sampai “z”. Serta, *case folding* mempunyai tujuan dalam menyeragamkan atau menyamakan kata yang sama namun memiliki penulisan yang berbeda. Hasil dari proses *case folding* dapat dilihat pada **Tabel 5.2** berikut.

**Tabel 5.2** Hasil *Case Folding*

Sebelum	Sesudah
<p><u>B</u>ahwa mereka terdakwa 1 <u>BURHAN</u> bin <u>ABDUL SAMAD</u> (alm) dan terdakwa 2 <u>SADAKA</u> bin <u>BASRI BABBAI</u> (alm) pada hari <u>Senin</u> tanggal 07 <u>Juni</u> 2021 sekitar jam 03.00 <u>Wib</u> atau setidaknya pada waktu lain dalam bulan <u>Juni</u> tahun 2021 bertempat di <u>Kantor Yayasan “BIAS” Ruko Jl. Tri tunggal Kav 2 Sorosutan Kec Umbulharjo Kota Yogyakarta</u>, atau setidaknya pada suatu tempat yang masih termasuk dalam daerah hukum <u>Pengadilan Negeri Yogyakarta</u>, para terdakwa telah</p>	<p><u>b</u>ahwa mereka terdakwa 1 <u>burhan</u> bin <u>abdul samad</u> (alm) dan terdakwa 2 <u>sadaka</u> bin <u>basri babbai</u> (alm) pada hari <u>senin</u> tanggal 07 juni 2021 sekitar jam 03.00 <u>wib</u> atau setidaknya pada waktu lain dalam bulan <u>juni</u> tahun 2021 bertempat di <u>kantor yayasan “bias” ruko jl. tri tunggal kav 2 sorosutan kec umbulharjo kota yogyakarta</u>, atau setidaknya pada suatu tempat yang masih termasuk dalam daerah hukum <u>pengadilan negeri yogyakarta</u>, para terdakwa telah mengambil barang sesuatu yang seluruhnya atau sebagian</p>

Sebelum	Sesudah
<p>mengambil barang sesuatu yang seluruhnya atau sebagian kepunyaan orang lain, dengan maksud untuk memiliki secara melawan hukum, pada waktu malam dalam sebuah rumah atau pekarangan tertutup yang ada rumahnya, yang dilakukan oleh orang yang adanya disitu tidak diketahui atau tidak dikehendaki oleh orang yang berhak, yang dilakukan oleh dua orang atau lebih bersama-sama yang untuk masuk ketempat melakukan kejahatan, atau untuk sampai pada barang yang diambilnya dilakukan dengan membongkar, memecah, atau memanjat atau dengan memakai anak kunci palsu, perintah palsu atau pakaian jabatan palsu. Perbuatan para terdakwa sebagaimana diatur dan diancam pidana dalam <b>P</b>asal 363 ayat (1) ke, 3, 4, dan 5 KUHP.</p>	<p>kepunyaan orang lain, dengan maksud untuk memiliki secara melawan hukum, pada waktu malam dalam sebuah rumah atau pekarangan tertutup yang ada rumahnya, yang dilakukan oleh orang yang adanya disitu tidak diketahui atau tidak dikehendaki oleh orang yang berhak, yang dilakukan oleh dua orang atau lebih bersama-sama yang untuk masuk ketempat melakukan kejahatan, atau untuk sampai pada barang yang diambilnya dilakukan dengan membongkar, memecah, atau memanjat atau dengan memakai anak kunci palsu, perintah palsu atau pakaian jabatan palsu. perbuatan para terdakwa sebagaimana diatur dan diancam pidana dalam <b>p</b>asal 363 ayat (1) ke, 3, 4, dan 5 <b>kuhp</b>.</p>

Pada **Tabel 5.2** merupakan data surat dakwaan sebelum dan sesudah dilakukan *case folding*. Dapat dilihat bahwa data surat dakwaan sebelum dilakukan proses *case folding* terdapat huruf besar/ kapital. Kemudian, setelah dilakukan proses *case folding* data surat dakwaan tidak terdapat lagi kata yang mengandung huruf besar/kapital.

### 5.2.2 *Remove Punctuation dan Numbers*

*Remove punctuation* merupakan tahapan yang bertujuan untuk menghapus atau membuang karakter berupa tanda baca seperti titik (.), koma (,), titik dua (:), tanda

hubung (-), tanda garis mirng dan sebagainya. sedangkan, untuk *remove numbers* digunakan untuk menghapus angka. Hasil dari proses *remove punctuation* dan *numbers* dapat dilihat pada **Tabel 5.3** berikut.

**Tabel 5.3** Hasil *Removing Punctuation* dan *Numbers*

Sebelum	Sesudah
<p>bahwa mereka terdakwa <u>1</u> burhan bin abdul samad (<u>alm</u>) dan terdakwa <u>2</u> sadaka bin basri babbai (<u>alm</u>) pada hari senin tanggal <u>07</u> juni <u>2021</u> sekitar jam <u>03.00</u> wib atau setidak_tidaknya pada waktu lain dalam bulan juni tahun <u>2021</u> bertempat di kantor yayasan "<u>bias</u>" ruko jl. tri tunggal kav 2 sorosutan kec umbulharjo kota yogyakarta. atau setidak_tidaknya pada suatu tempat yang masih termasuk dalam daerah hukum pengadilan negeri yogyakarta. para terdakwa telah mengambil barang sesuatu yang seluruhnya atau sebagian kepunyaan orang lain. dengan maksud untuk memiliki secara melawan hukum. pada waktu malam dalam sebuah rumah atau pekarangan tertutup yang ada rumahnya. yang dilakukan oleh orang yang adanya disitu tidak diketahui atau tidak dikehendaki oleh orang yang berhak. yang dilakukan oleh dua orang atau lebih bersama-sama yang untuk masuk ketempat melakukan kejahatan. atau untuk sampai pada barang yang diambilnya dilakukan dengan membongkar,</p>	<p>bahwa mereka terdakwa burhan bin abdul samad alm dan terdakwa sadaka bin basri babbai alm pada hari senin tanggal juni sekitar jam wib atau setidak tidaknya pada waktu lain dalam bulan juni tahun bertempat di kantor yayasan bias ruko jl tri tunggal kav sorosutan kec umbulharjo kota yogyakarta atau setidak tidaknya pada suatu tempat yang masih termasuk dalam daerah hukum pengadilan negeri yogyakarta para terdakwa telah mengambil barang sesuatu yang seluruhnya atau sebagian kepunyaan orang lain dengan maksud untuk memiliki secara melawan hukum pada waktu malam dalam sebuah rumah atau pekarangan tertutup yang ada rumahnya yang dilakukan oleh orang yang adanya disitu tidak diketahui atau tidak dikehendaki oleh orang yang berhak yang dilakukan oleh dua orang atau lebih bersama sama yang untuk masuk ketempat melakukan kejahatan atau untuk sampai pada barang yang diambilnya dilakukan dengan membongkar memecah atau memanjat atau dengan memakai anak kunci palsu</p>

Sebelum	Sesudah
memecah, atau memanjat atau dengan memakai anak kunci palsu, perintah palsu atau pakaian jabatan palsu, perbuatan para terdakwa sebagaimana diatur dan diancam pidana dalam pasal <u>363</u> ayat <u>(1)</u> ke, <u>3</u> , <u>4</u> , dan <u>5</u> kuhp.	perintah palsu atau pakaian jabatan palsu perbuatan para terdakwa sebagaimana diatur dan diancam pidana dalam pasal ayat ke dan kuhp

Pada **Tabel 5.3** merupakan data surat dakwaan sebelum dan sesudah dilakukan proses *remove punctuation* dan *numbers*. Dapat dilihat bahwa data surat dakwaan sebelum dilakukan proses *remove punctuation* dan *numbers* terdapat tanda baca dan angka. Kemudian, setelah dilakukan proses *remove punctuation* dan *numbers* data surat dakwaan tidak terdapat lagi kata yang mengandung tanda baca dan angka.

### 5.2.3 Stopword

*Stopword* merupakan tahapan yang bertujuan untuk menghapus kata-kata yang sering muncul namun tidak penting atau tidak memiliki makna pasti dalam data. Contoh dalam proses *stopword* yaitu “dan”, “di”, “ke”, “yang”, “dengan”, “atau” dan lain-lain. Hasil dari proses *stopword* dapat dilihat pada **Tabel 5.4** berikut.

**Tabel 5.4** Hasil *Stopword*

Sebelum	Sesudah
<u>bahwa mereka terdakwa</u> burhan <u>bin</u> <u>abdul samad alm dan terdakwa</u> sadaka <u>bin</u> basri babbai <u>alm pada hari senin</u> <u>tanggal juni sekitar jam wib atau</u> <u>setidak tidaknya pada waktu lain dalam</u> <u>bulan juni tahun</u> bertempat <u>di</u> kantor yayasan bias ruko <u>jl</u> tri tunggal kav sorosutan <u>kec</u> umbulharjo <u>kota</u> <u>yogyakarta</u> atau setidaknya <u>tidaknya pada</u> <u>suatu tempat yang masih termasuk</u> <u>dalam daerah</u> hukum <u>pengadilan negeri</u>	burhan abdul samad sadaka basri babbai bertempat kantor yayasan bias ruko tri tunggal kav sorosutan umbulharjo hukum mengambil barang kepunyaan orang maksud memiliki melawan hukum malam rumah pekarangan tertutup rumahnya orang disitu dikehendaki orang berhak orang masuk ketempat kejahatan barang diambilnya membongkar memecah memanjat memakai anak kunci palsu perintah

Sebelum	Sesudah
<p>yogyakarta para terdakwa telah mengambil barang sesuatu yang seluruhnya atau sebagian kepunyaan orang lain dengan maksud untuk memiliki secara melawan hukum pada waktu malam dalam sebuah rumah atau pekarangan tertutup yang ada rumahnya yang dilakukan oleh orang yang adanya disitu tidak diketahui atau tidak dikehendaki oleh orang yang berhak yang dilakukan oleh dua orang atau lebih bersama sama yang untuk masuk ketempat melakukan kejahatan atau untuk sampai pada barang yang diambilnya dilakukan dengan membongkar memecah atau mem anjat atau dengan memakai anak kunci palsu perintah palsu atau pakaian jabatan palsu perbuatan para terdakwa sebagaimana diatur dan diancam pidana dalam pasal ayat ke dan kuhp</p>	<p>palsu pakaian jabatan palsu perbuatan diatur diancam pidana</p>

Pada **Tabel 5.4** merupakan data surat dakwaan sebelum dan sesudah dilakukan proses *stopword*. Dapat dilihat bahwa data surat dakwaan sebelum dilakukan proses *stopword* terdapat kata-kata yang sering muncul namun tidak penting atau tidak memiliki makna. Kemudian, setelah dilakukan proses *stopword* data surat dakwaan tidak terdapat lagi kata-kata tersebut.

#### 5.2.4 Stemming

*Stemming* merupakan tahapan yang bertujuan untuk menghilangkan atau menghapus semua kata yang mempunyai imbuhan diawal atau akhir dan

mengubahnya menjadi kata dasar. Hasil dari proses *stemming* dapat dilihat pada **Tabel 5.5** berikut.

**Tabel 5.5** Hasil *Stemming*

Sebelum	Sesudah
burhan abdul samad sadaka basri babbai <u>bertempat</u> kantor yayasan bias ruko tri tunggal kav sorosutan umbulharjo hukum <u>mengambil</u> barang <u>kepunyaan</u> orang maksud <u>memiliki</u> <u>melawan</u> hukum malam rumah <u>pekarangan</u> <u>tertutup</u> <u>rumahnya</u> orang <u>disitu</u> <u>dikehendaki</u> orang <u>berhak</u> orang masuk <u>ketempat</u> <u>kejahatan</u> barang <u>diambilnya</u> <u>membongkar</u> <u>memecah</u> <u>memanjat</u> <u>memakai</u> anak kunci palsu perintah palsu <u>pakaian</u> <u>jabatan</u> palsu <u>perbuatan</u> <u>diatur</u> <u>diancam</u> pidana	burhan abdul samad sadaka basri babbai <u>tempat</u> kantor yayasan bias ruko tri tunggal kav sorosutan umbulharjo hukum <u>ambil</u> barang <u>punya</u> orang maksud <u>milik</u> <u>lawan</u> hukum malam rumah <u>pekarang</u> <u>tutup</u> <u>rumah</u> orang <u>situ</u> <u>hendak</u> orang <u>hak</u> orang masuk <u>tempat</u> <u>jahat</u> barang <u>ambil</u> <u>bongkar</u> <u>pecah</u> <u>panjang</u> <u>pakai</u> anak kunci palsu perintah palsu <u>pakai</u> <u>jabat</u> palsu <u>buat</u> <u>atur</u> <u>ancam</u> pidana

Pada **Tabel 5.5** merupakan data surat dakwaan sebelum dan sesudah dilakukan proses *stemming*. Dapat dilihat bahwa data surat dakwaan sebelum dilakukan proses *stemming* terdapat kata seperti “bertempat” dan “membongkar”. Kemudian, setelah dilakukan proses *stemming* kata tersebut berubah menjadi “tempat” dan “bongkar”.

### 5.2.5 Tokenizing

*Tokenizing* merupakan tahapan yang bertujuan untuk membagi suatu kata yang berbentuk kalimat dalam dokumen menjadi unit yang lebih kecil. Tahapan *tokenizing* sangat diperlukan untuk tahapan selanjutnya, karena dapat memudahkan dalam dalam hal pembobotan kata pada tahapan *TF-IDF*. Hasil dari proses *tokenizing* dapat dilihat pada **Tabel 5.6** berikut.

**Tabel 5.6** Hasil *Tokenizing*

Sebelum	Sesudah
burhan abdul samad sadaka basri babbai tempat kantor yayasan bias ruko tri tunggal kav sorosutan umbulharjo hukum ambil barang punya orang maksud milik lawan hukum malam rumah pekarang tutup rumah orang situ hendak orang hak orang masuk tempat jahat barang ambil bongkar pecah panjat pakai anak kunci palsu perintah palsu pakai jabat palsu buat atur ancam pidana	['burhan', 'abdul', 'samad', 'sadaka', 'basri', 'babbai', 'tempat', 'kantor', 'yayasan', 'bias', 'ruko', 'tri', 'tunggal', 'kav', 'sorosutan', 'umbulharjo', 'hukum', 'ambil', 'barang', 'punya', 'orang', 'maksud', 'milik', 'lawan', 'hukum', 'malam', 'rumah', 'pekarang', 'tutup', 'rumah', 'orang', 'situ', 'hendak', 'orang', 'hak', 'orang', 'masuk', 'tempat', 'jahat', 'barang', 'ambil', 'bongkar', 'pecah', 'panjat', 'pakai', 'anak', 'kunci', 'palsu', 'perintah', 'palsu', 'pakai', 'jabat', 'palsu', 'buat', 'atur', 'ancam', 'pidana']

Pada **Tabel 5.6** merupakan data surat dakwaan sebelum dan sesudah dilakukan proses *tokenizing*. Dapat dilihat bahwa data surat dakwaan sebelum dilakukan proses *tokenizing* terdapat kata yang berbentuk kalimat belum dipisahkan. Kemudian, setelah dilakukan proses *tokenizing* kata yang berbentuk kalimat tersebut telah terpisah menjadi unit yang lebih kecil.

### 5.3. *Term Frequency-Invers Document Frequency (TF-IDF)*

*Term Frequency-Invers Document Frequency (TF-IDF)* merupakan tahap untuk melakukan pembobotan setelah melakukan *preprocessing* data. Pada tahapan *TF-IDF* akan dilakukan perubahan data yang masih berbentuk kata menjadi numerik atau angka. *TF-IDF* merupakan gabungan dari dua metode yaitu *TF* dan *IDF*. Metode *TF* merupakan metode yang bertujuan untuk menghitung frekuensi dari setiap kemunculan kata dalam dokumen. sedangkan, metode *IDF* merupakan metode yang bertujuan untuk menghitung frekuensi dokumen yang memuat sebuah kata. Dalam penulisan ini, penulis mengambil beberapa sampel kata yaitu curi, lawan, luka, narkoba, senjata dan serang. Hasil dari tahapan *TF-IDF* dapat dilihat pada **Tabel 5.7** berikut.



**Tabel 5.7** Hasil Pembobotan *TF-IDF* dengan *Phyton*

<i>Doc</i>	Curi	...	Lawan	Luka	...	Narkoba	...	Senjata	Serang
13	8,7039	...	4,1224	0	...	0	...	0	0
:	:	...	:	:	...	:	...	:	:
23	0	...	0	78,8727	...	0	...	0	6,2688
24	0	...	0	17,5272	...	0	...	0	6,2688
25	0	...	0	0	...	6,6279	...	0	0
:	:	:	:	:	...	:	...	:	:
54	4,3519	...	0	0	...	0	...	0	0
55	34,8157	...	2,0612	0	...	0	...	0	0
:	:	:	:	:	...	:	...	:	:
77	0	...	0	0	...	0	...	13,8606	0

Pada **Tabel 5.7** merupakan hasil dari perhitungan *TF-IDF* menggunakan *software Phyton*. Kemudian, untuk perhitungan manual dalam *TF-IDF* dilakukan dengan menghitung *Term Frequency (TF)* seperti pada **Tabel 5.8** berikut.

**Tabel 5.8** Hasil Pembobotan *TF* dengan *Phyton*

<i>Doc</i>	Curi	...	Lawan	Luka	...	Narkoba	...	Senjata	Serang
13	2	...	2	0	...	0	...	0	0
:	:	...	:	:	...	:	...	:	:
23	0	...	0	18	...	0	...	0	1
24	0	...	0	4	...	0	...	0	1
25	0	...	0	0	...	2	...	0	0
:	:	:	:	:	...	:	...	:	:
54	1	...	0	0	...	0	...	0	0
55	8	...	1	0	...	0	...	0	0
:	:	:	:	:	...	:	...	:	:
77	0	...	0	0	...	0	...	3	0

Pada **Tabel 5.8** merupakan hasil pembobotan *TF* menggunakan *Phyton* untuk data surat dakwaan dan dapat dilihat banyaknya kata yang muncul pada masing-masing dokumen berdasarkan kata sampel yang penulis gunakan. Pada tabel *TF*

untuk perhitungan yang menghasilkan nilai 0 menunjukkan bahwa tidak terdapat kemunculan kata (*term*) dalam dokumen tersebut. Kemudian untuk perhitungan *TF* yang menghasilkan nilai selain 0 seperti 1, 2, 3 dan sebagainya menunjukkan bahwa terdapat kemunculan kata (*term*) dalam dokumen tersebut. Misalkan untuk kata “curi” pada dokumen ke-13 menghasilkan nilai atau muncul sebanyak 2 kali, maka nilai *TF* yang digunakan adalah 2. Setelah mendapatkan pembobotan *TF* dilanjutkan mencari perhitungan *DF* secara manual, seperti berikut.

**Tabel 5.9** Hasil Pembobotan *DF* Perhitungan Manual

Kata	TF		DF	Total Doc (D)
	Dokumen 54	Dokumen 55		
Curi	1	8	34	971
Lawan	0	1	336	

Pada **Tabel 5.9** merupakan nilai *DF* dari kata “curi” dan “lawan” yang terdapat dalam dokumen ke-54 dan ke-55 dari total seluruh data dokumen dakwaan yaitu 971. Kata “curi” pada keseluruhan dokumen muncul sebanyak 34 kali dan kata “lawan” pada keseluruhan dokumen muncul sebanyak 336 kali. Kemudian melakukan perhitungan bobot *TF-IDF* menggunakan hasil dari *TF* dan *DF* yang telah didapatkan sebelumnya pada **Tabel 5.8** dan **Tabel 5.9**, seperti berikut.

**Tabel 5.10** Hasil Pembobotan *TD-IDF* Perhitungan Manual

Kata	DF	$\frac{D}{DF}$	IDF $(\ln(\frac{D}{DF})+1)$	Dokumen 54		Dokumen 55	
				TF	(TF x IDF)	TF	(TF x IDF)
Curi	34	28,5588	4,3519	1	4,3519	8	34,8157
Lawan	336	2,8899	2,0612	0	0	1	2,0612

Pada **Tabel 5.10** merupakan tahapan perhitungan dari manual dari *TF-IDF* untuk kata “curi” dan “lawan” pada dokumen ke-54 dan ke-55. Tahapan awal dilakukan perhitungan  $\frac{D}{DF}$  untuk mencari nilai *IDF*, dengan *D* merupakan jumlah dari seluruh dokumen yaitu 971 dan *DF* pada perhitungan ini menggunakan nilai dari kata “curi” muncul sebanyak 34 kali dan “lawan” muncul sebanyak “336” kali. Kemudian melakukan perhitungan *IDF* dengan mencari nilai *ln* dari nilai *D* dibagi dengan *DF*, lalu ditambah 1 sehingga didapatlah nilai *IDF* dari kata “curi” yaitu 4,3519 dan kata “lawan” yaitu 2,0612.

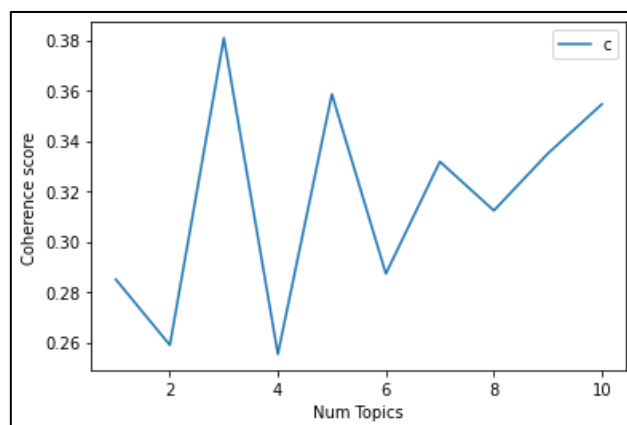
Setelah itu, melakukan perhitungan *TF-IDF* dengan mengkalikan nilai *TF* dan *IDF*. Sehingga didapatkan nilai *TD-IDF* untuk kata “curi” pada dokumen ke-54 yaitu 4,3519 dan dokumen ke-55 yaitu 34,8157. Sedangkan nilai *TF-IDF* dari kata “lawan” pada dokumen ke-54 yaitu 0 dan dokumen ke-55 yaitu 2,0612.

#### 5.4. Hasil *Topic Modelling*

Tahapan setelah melakukan pembobotan *TF-IDF* yaitu melakukan *topic modelling* dengan metode *Latent Dirichlet Allocation (LDA)* yang bertujuan untuk melihat topik yang banyak terjadi dan dibahas dalam suatu data atau dokumen, dimana pada penelitian ini analisis *topic modelling* didasarkan pada data surat dakwaan dari klasifikasi perkara kasus tindak pidana yang terdiri dari narkoba, pencurian dan kesehatan di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.

##### 5.4.1 Narkoba

Pada tahap *topic modelling* terlebih dahulu ditentukan jumlah topik menggunakan *coherence score*. *Coherence score* merupakan ukuran yang digunakan sebagai evaluasi dari *topic modelling* yang divisualisasikan dalam bentuk grafik. Semakin besar *coherence score*, maka model yang dihasilkan akan semakin baik. Berikut adalah grafik *coherence score* dengan menggunakan limit topik yaitu 1-11 dari data klasifikasi perkara narkoba di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.



**Gambar 5.4** Grafik *Coherence Score* Narkoba

Pada **Gambar 5.4** merupakan visualisasi dari grafik *coherence score* untuk klasifikasi perkara narkoba, dengan nilai tertinggi terdapat pada topik ke-3. Sehingga jumlah topik yang terbentuk dari klasifikasi perkara narkoba di Pengadilan Negeri Yogyakarta pada tahun 2019-2021 adalah sebanyak 3. Besarnya

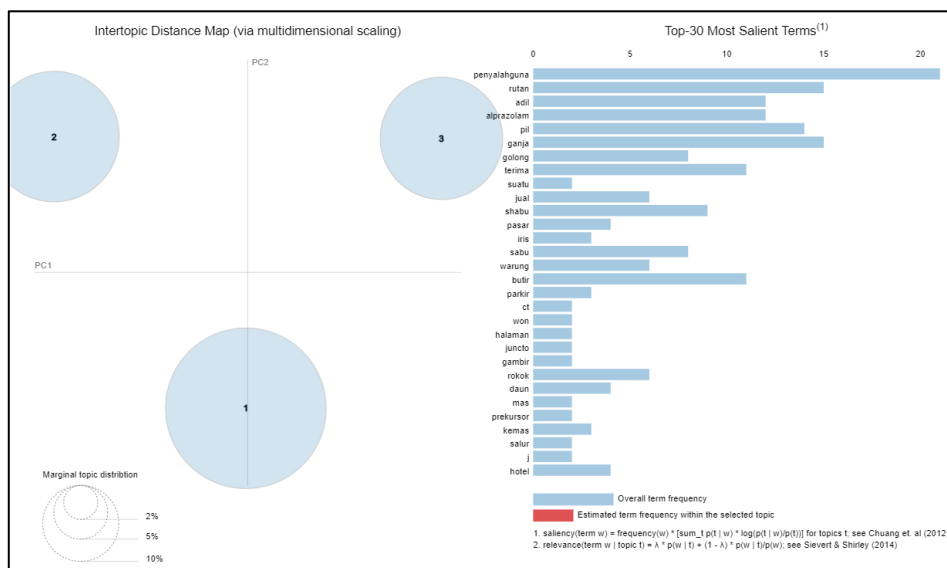
*coherence score* atau nilai koherensi dari topik berdasarkan grafik pada **Gambar 5.4** dapat dilihat pada **Tabel 5.11** berikut.

**Tabel 5.11** Nilai Koherensi Topik Narkotika

<i>Num Topics</i>	<i>Coherence Value</i>	<i>Num Topics</i>	<i>Coherence Value</i>
1	0,2850	6	0,2873
2	0,2589	7	0,3318
3	0,3809	8	0,3124
4	0,2553	9	0,3350
5	0,3586	10	0,3546

Berdasarkan nilai pada **Tabel 5.11** dapat dilihat bahwa *num topics* 3 merupakan *coherence value* tertinggi dengan nilai sebesar 0,3809. Hal ini menunjukkan bahwa jumlah topik yang terbentuk terkait klasifikasi perkara narkotika untuk mengetahui kasus tindak pidana yang sering terjadi di Pengadilan Negeri Yogyakarta adalah sebanyak 3 topik. Hasil tersebut sama dan sesuai dengan visualisasi grafik *coherence score* pada **Gambar 5.4**. Sehingga jumlah 3 topik yang dihasilkan tersebut akan dijadikan acuan dalam pembuatan model.

Selanjutnya, peneliti melakukan analisis *topic modelling* menggunakan visualisasi dari *pyLDAvis* untuk mengetahui keterkaitan antar kata yang sering muncul pada suatu topik. Berikut adalah visualisasi dari 3 topik menggunakan *pyLDAvis*.



**Gambar 5.5** Visualisasi 3 Topik Perkara Narkotika

Pada visualisasi *pyLDAvis* menunjukkan diagram terbagi menjadi 2 panel, yaitu panel kiri dan panel kanan. Pada panel kiri merupakan pemetaan jarak antar topik (*intertopic distance map*), divisualisasikan dengan lingkaran-lingkaran berisikan angka yang menunjukkan *num topics*. Besar kecilnya suatu lingkaran menggambarkan seberapa penting dan sering topik tersebut muncul dalam dokumen surat dakwaan. Selain itu dapat dilihat pada visualisasi *intertopic distance map* terbagi menjadi 2 sumbu yaitu *Principal Component 1 (PC1)* dan *Principal Component 2 (PC2)*, sehingga terbentuk kuadran yang memiliki makna bahwa lingkaran topik yang berdekatan memiliki hubungan yang kuat dan lingkaran topik yang berjauhan memiliki hubungan yang lemah. Berikut adalah nilai *principal component* dari 3 topik yang digunakan.

**Tabel 5.12** Hasil *Principal Component*

<i>Num Topics</i>	<b>PC1</b>	<b>PC2</b>
1	-0,0001	-0,0270
2	-0,0288	0,0136
3	0,0290	0,0133

Berdasarkan **Tabel 5.12** hasil *Principal Component* dapat dilihat bahwa nilai *PC1* merupakan titik koordinat pada sumbu *X* dan nilai *PC2* merupakan titik koordinat pada sumbu *Y*. Sehingga dengan melihat nilai *PC1* dan *PC2* dapat diketahui bahwa topik ke-1 berada di kuadran III ditandai dengan sumbu *X* bernilai negatif dan sumbu *Y* bernilai negatif mempunyai arti bahwa representasi kata pada model rendah dan mempunyai variasi kata yang rendah, namun berkaitan dengan topik tertentu. Topik ke-2 berada di kuadran II ditandai dengan sumbu *X* bernilai negatif dan sumbu *Y* bernilai positif mempunyai arti bahwa representasi kata pada model rendah dan mempunyai variasi kata yang tinggi, namun lebih berkaitan dengan topik tambahan. Topik ke-3 berada di kuadran I ditandai dengan sumbu *X* bernilai positif dan sumbu *Y* bernilai positif mempunyai arti bahwa representasi kata pada model tinggi, dan mempunyai variasi kata yang tinggi, sehingga berkaitan dengan topik utama.

Pada panel kanan merupakan 30 kata yang sering muncul pada topik dari keseluruhan dokumen, divisualisasikan dengan diagram batang berwarna biru dan merah. Diagram batang berwarna biru merupakan indikasi *term frequency* yang

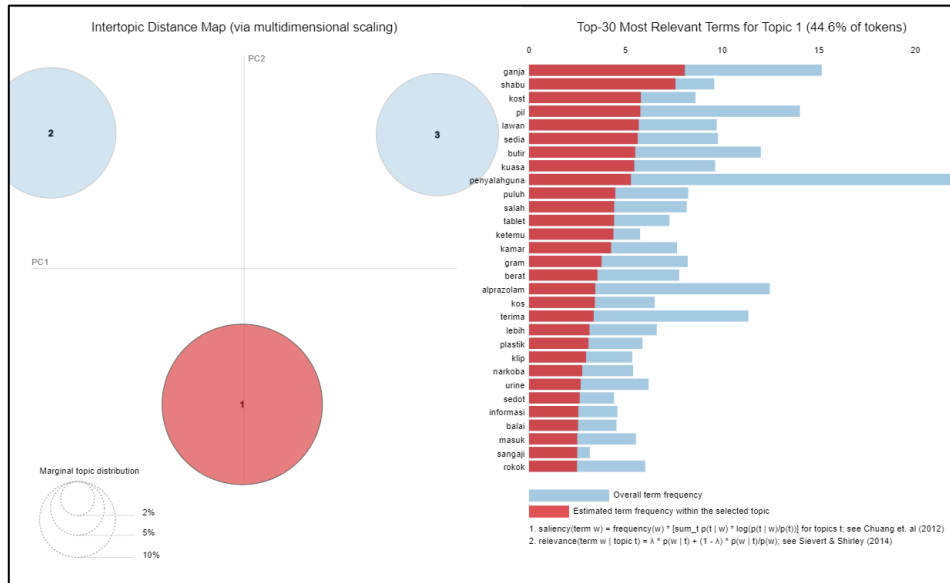
sering muncul dalam topik. Diagram batang berwarna merah merupakan estimasi *term frequency* pada suatu topik tertentu.

Berikut merupakan visualisasi berupa *word cloud* dari 3 topik untuk menampilkan frekuensi kata-kata yang sering muncul pada teks atau dokumen agar mudah dipahami oleh pembaca dari data surat dakwaan untuk klasifikasi perkara narkoba di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.



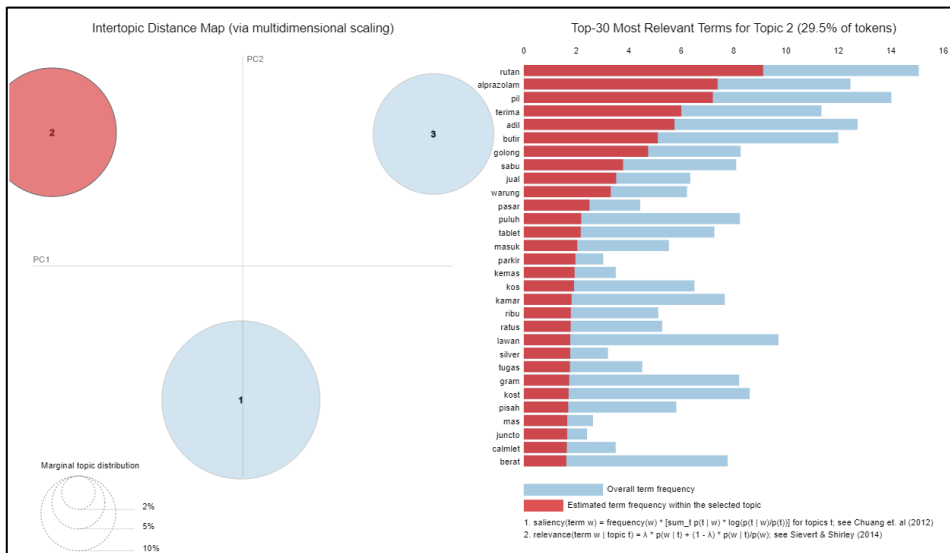
**Gambar 5.6** *Word Cloud* Perkara Narkotika

Pada **Gambar 5.6** merupakan visualisasi *word cloud* dari 3 topik. Semakin banyak frekuensi kata pada data *set*, maka akan semakin besar huruf yang akan ditampilkan pada *word cloud*. Sebaliknya, semakin sedikit frekuensi kata pada data *set* maka semakin kecil huruf yang akan ditampilkan pada *word cloud*. Berdasarkan **Gambar 5.6** terlihat bahwa frekuensi kata yang ditampilkan paling besar terdapat pada kata “rutan”, “alprazolam” dan “pil”. Hasil tersebut sesuai dengan visualisasi *pyLDavis* yang terdapat pada **Gambar 5.5**. Kemudian, melakukan visualisasi *pyLDavis* untuk topik 1 seperti berikut.



**Gambar 5.7** Visualisasi Topik 1

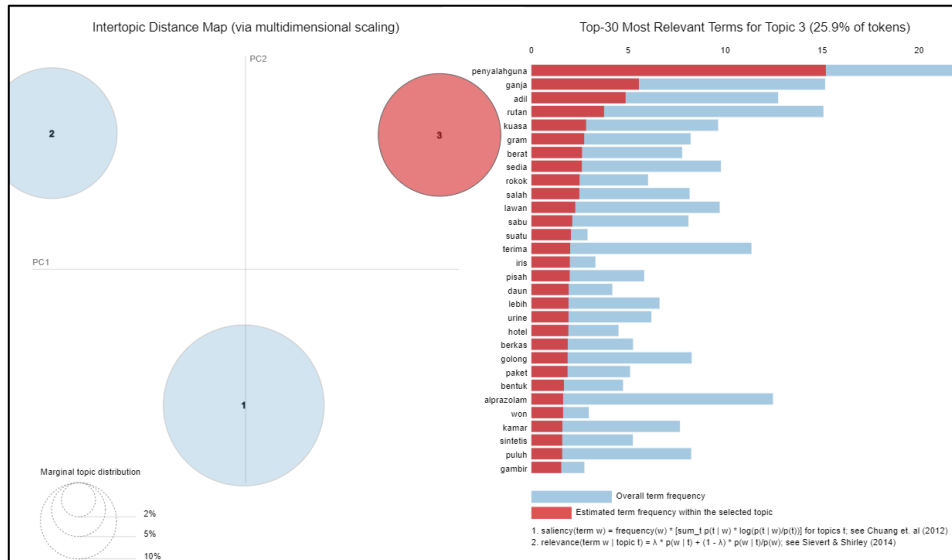
Pada **Gambar 5.7** visualisasi dari topik 1 ditunjukkan dengan lingkaran berwarna merah dan estimasi *term frequency* topik 1 ditunjukkan dengan diagram batang berwarna merah. *Marginal topic distribution* pada topik 1 mempunyai nilai sebesar 44,6% yang artinya tingkat pentingnya topik 1 terhadap dokumen memuat 44,6% dari keseluruhan topik dalam dokumen. Pada diagram batang berwarna merah memperlihatkan 30 frekuensi kata yang relevan dengan topik 1. Dapat dilihat bahwa 3 kata yang sering muncul pada topik 1 adalah “ganja”, “shabu” dan “kost”.



**Gambar 5.8** Visualisasi Topik 2

Pada **Gambar 5.8** visualisasi dari topik 2 ditunjukkan dengan lingkaran berwarna merah dan estimasi *term frequency* topik 2 ditunjukkan dengan diagram batang berwarna merah. *Marginal topic distribution* pada topik 2 mempunyai nilai

sebesar 29,5% yang artinya tingkat pentingnya topik 2 terhadap dokumen memuat 29,5% dari keseluruhan topik dalam dokumen. Pada diagram batang berwarna merah memperlihatkan 30 frekuensi kata yang relevan dengan topik 2. Dapat dilihat bahwa 3 kata yang sering muncul pada topik 2 adalah “rutan”, “alprazolam” dan “pil”.



**Gambar 5.9** Visualisasi Topik 3

Pada **Gambar 5.9** visualisasi dari topik 3 ditunjukkan dengan lingkaran berwarna merah dan estimasi *term frequency* topik 3 ditunjukkan dengan diagram batang berwarna merah. *Marginal topic distribution* pada topik 3 mempunyai nilai sebesar 25,9% yang artinya tingkat pentingnya topik 3 terhadap dokumen memuat 25,9% dari keseluruhan topik dalam dokumen. Pada diagram batang berwarna merah memperlihatkan 30 frekuensi kata yang relevan dengan topik 3. Dapat dilihat bahwa 3 kata yang sering muncul pada topik 3 adalah “penyalahguna”, “ganja” dan “adi”.

Setelah melakukan visualisasi 3 topik menggunakan *pyLDAvis*, maka dapat disimpulkan kasus tindak pidana yang sering terjadi dan dibahas di Pengadilan Negeri Yogyakarta berdasarkan data dokumen surat dakwaan pada tahun 2019-2021 dengan membuat pemodelan topik. Pemodelan topik dilakukan menggunakan kata-kata atau *term frequency* yang sering muncul dengan cara menyambungkan kata yang relevan dan berkaitan satu sama lain. Berikut adalah pemodelan 3 topik dari data klasifikasi perkara narkoba di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.



**Tabel 5.13** Hasil Pemodelan 3 Topik Narkotika

Topik	Pemodelan Topik
1	0,013*“ganja” + 0,012*“shabu” + 0,009*“kost” + 0,009*“pil” + 0,009*“lawan” + 0,009*“sedia” + 0,009*“butir” + 0,009*“kuasa” + 0,008*“penyalahguna” + 0,007*“puluh”
2	0,022*“rutan” + 0,018*“alprazolam” + 0,017*“pil” + 0,014*“terima” + 0,014*“adil” + 0,014*“butir” + 0,012*“golong” + 0,011*“sabu” + 0,009*“jual” + 0,007*“warung”
3	0,041*“penyalahguna” + 0,015*“ganja” + 0,013*“adil” + 0,010*“rutan” + 0,008*“kuasa” + 0,007*“gram” + 0,007*“berat” + 0,007*“sedia” + 0,007*“rokok” + 0,007*“salah”

Berdasarkan **Tabel 5.13** hasil pemodelan 3 topik menggunakan metode *Latent Dirichlet Allocation (LDA)* didapatkan kesimpulan bahwa pada klasifikasi perkara narkotika banyak membahas tentang jenis kasus yang terjadi, zat narkotika yang digunakan dan tempat terjadinya peristiwa. Berikut adalah kesimpulan dari masing-masing topik pada klasifikasi perkara narkotika.

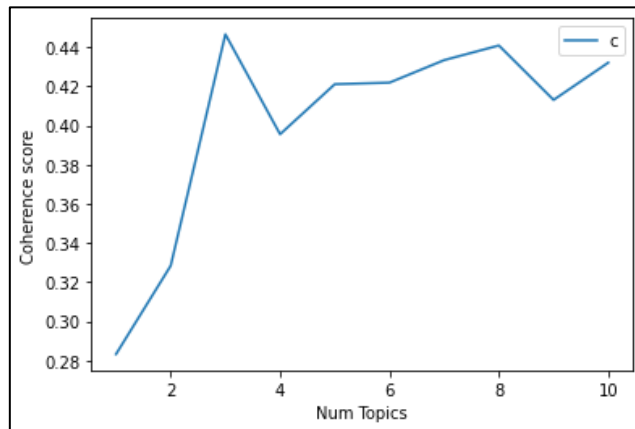
**Tabel 5.14** Kesimpulan Pemodelan 3 Topik Narkotika

Topik	Jenis Kasus	Zat	Tempat Terjadi
1	Menyediakan	Sabu	Kontrakan
2	Menjual	Alprazolam	Rutan
3	Menjual	Ganja	Rutan

Pada **Tabel 5.14** kesimpulan dari masing-masing pemodelan 3 topik klasifikasi perkara narkotika, dapat diketahui topik 1 banyak membahas tentang kasus menyediakan zat narkotika yaitu sabu dengan tempat terjadinya peristiwa di kontrakan. Topik 2 banyak membahas tentang kasus menjual zat narkotika yaitu alprazolam dengan tempat terjadinya peristiwa di rutan. Topik 3 banyak membahas tentang kasus menjual zat narkotika yaitu ganja dengan tempat terjadi peristiwa di rutan.

### 5.4.2 Pencurian

Pada tahap *topic modelling* terlebih dahulu ditentukan jumlah topik menggunakan *coherence score*. *Coherence score* merupakan ukuran yang digunakan sebagai evaluasi dari *topic modelling* yang divisualisasikan dalam bentuk grafik. Semakin besar *coherence score*, maka model yang dihasilkan akan semakin baik. Berikut adalah grafik *coherence score* dengan menggunakan limit topik yaitu 1-11 dari data klasifikasi perkara pencurian di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.



**Gambar 5.10** Grafik *Coherence Score* Pencurian

Pada **Gambar 5.10** merupakan visualisasi dari grafik *coherence score* untuk klasifikasi perkara pencurian, dengan nilai tertinggi terdapat pada topik ke-3. Sehingga jumlah topik yang terbentuk dari klasifikasi perkara pencurian di Pengadilan Negeri Yogyakarta pada tahun 2019-2021 adalah sebanyak 3. Besarnya *coherence score* atau nilai koherensi dari topik berdasarkan grafik pada **Gambar 5.10** dapat dilihat pada **Tabel 5.15** berikut.

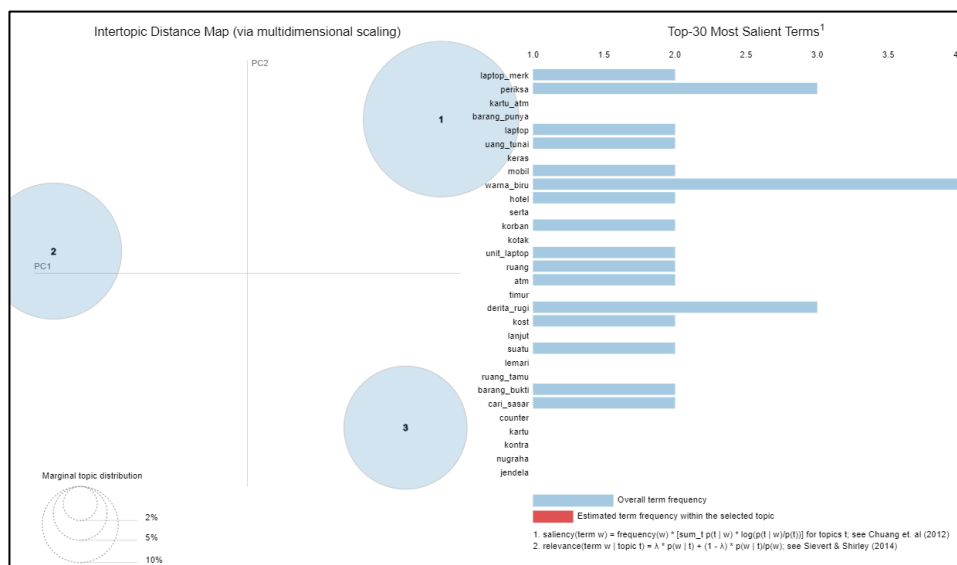
**Tabel 5.15** Nilai Koherensi Topik Pencurian

<i>Num Topics</i>	<i>Coherence Value</i>	<i>Num Topics</i>	<i>Coherence Value</i>
1	0,2831	6	0,4219
2	0,3283	7	0,4334
3	0,4466	8	0,4409
4	0,3955	9	0,4131
5	0,4211	10	0,4321

Berdasarkan nilai pada **Tabel 5.15** dapat dilihat bahwa *num topics* 3 merupakan *coherence value* tertinggi dengan nilai sebesar 0,4466. Hal ini menunjukkan bahwa

jumlah topik yang terbentuk terkait klasifikasi perkara pencurian untuk mengetahui kasus tindak pidana yang sering terjadi di Pengadilan Negeri Yogyakarta adalah sebanyak 3 topik. Hasil tersebut sama dan sesuai dengan visualisasi grafik *coherence score* pada **Gambar 5.10**. Sehingga jumlah 3 topik yang dihasilkan tersebut akan dijadikan acuan dalam pembuatan model.

Selanjutnya, peneliti melakukan analisis *topic modelling* menggunakan visualisasi dari *pyLDAvis* untuk mengetahui keterkaitan antar kata yang sering muncul pada suatu topik. Berikut adalah visualisasi dari 3 topik menggunakan *pyLDAvis*.



**Gambar 5.11** Visualisasi 3 Topik Perkara Pencurian

Pada visualisasi *pyLDAvis* menunjukkan diagram terbagi menjadi 2 panel, yaitu panel kiri dan panel kanan. Pada panel kiri merupakan pemetaan jarak antar topik (*intertopic distance map*), divisualisasikan dengan lingkaran-lingkaran berisikan angka yang menunjukkan *num topics*. Besar kecilnya suatu lingkaran menggambarkan seberapa penting dan sering topik tersebut muncul dalam dokumen surat dakwaan. Selain itu dapat dilihat pada visualisasi *intertopic distance map* terbagi menjadi 2 sumbu yaitu *Principal Component 1 (PC1)* dan *Principal Component 2 (PC2)*, sehingga terbentuk kuadran yang memiliki makna bahwa lingkaran topik yang berdekatan memiliki hubungan yang kuat dan lingkaran topik yang berjauhan memiliki hubungan yang lemah. Berikut adalah nilai *principal component* dari 3 topik yang digunakan.

**Tabel 5.16** Hasil *Principial Component*

<i>Num Topics</i>	<i>PC1</i>	<i>PC2</i>
1	0,0169	0,0176
2	-0,0296	0,0017
3	0,0126	-0,0194

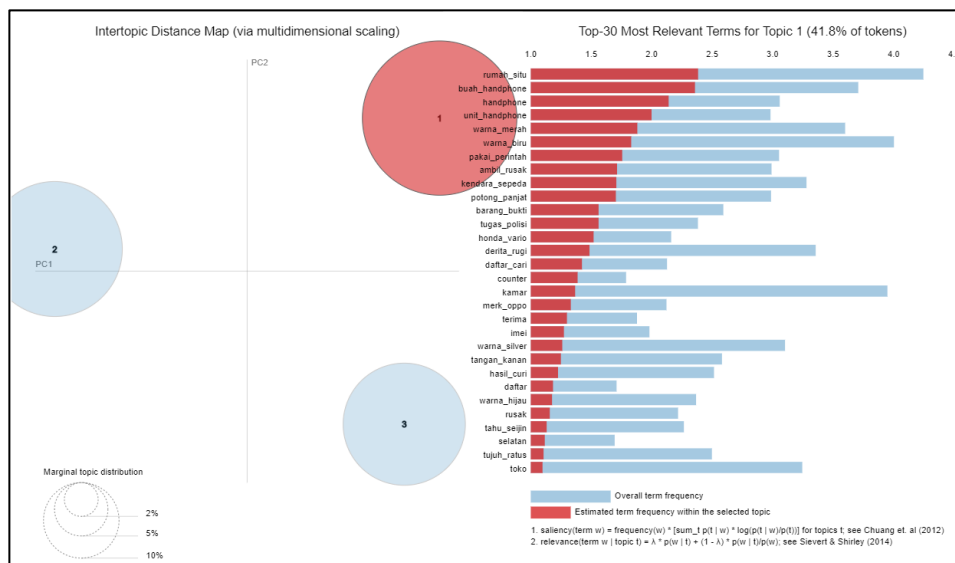
Berdasarkan **Tabel 5.16** hasil *Principal Component* dapat dilihat bahwa nilai *PC1* merupakan titik koordinat pada sumbu *X* dan nilai *PC2* merupakan titik kordinat pada sumbu *Y*. Sehingga dengan melihat nilai *PC1* dan *PC2* dapat diketahui bahwa topik ke-1 berada di kuadran I ditandai dengan sumbu *X* bernilai positif dan sumbu *Y* bernilai positif mempunyai arti bahwa representasi kata pada model tinggi, dan mempunyai variasi kata yang tinggi, sehingga berkaitan dengan topik utama. Topik ke-2 berada di kuadran II ditandai dengan sumbu *X* bernilai negatif dan sumbu *Y* bernilai positif mempunyai arti bahwa representasi kata pada model rendah dan mempunyai variasi kata yang tinggi, namun lebih berkaitan dengan topik tambahan. Topik ke-3 berada di kuadran IV ditandai dengan sumbu *X* bernilai positif dan sumbu *Y* bernilai negatif mempunyai arti bahwa representasi kata pada model tinggi dan mempunyai variasi kata yang rendah, sehingga kurang berkaitan dengan topik utama.

Berikut merupakan visualisasi berupa *word cloud* dari 3 topik untuk menampilkan frekuensi kata-kata yang sering muncul pada teks atau dokumen agar mudah dipahami oleh pembaca dari data surat dakwaan untuk klasifikasi perkara pencurian di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.



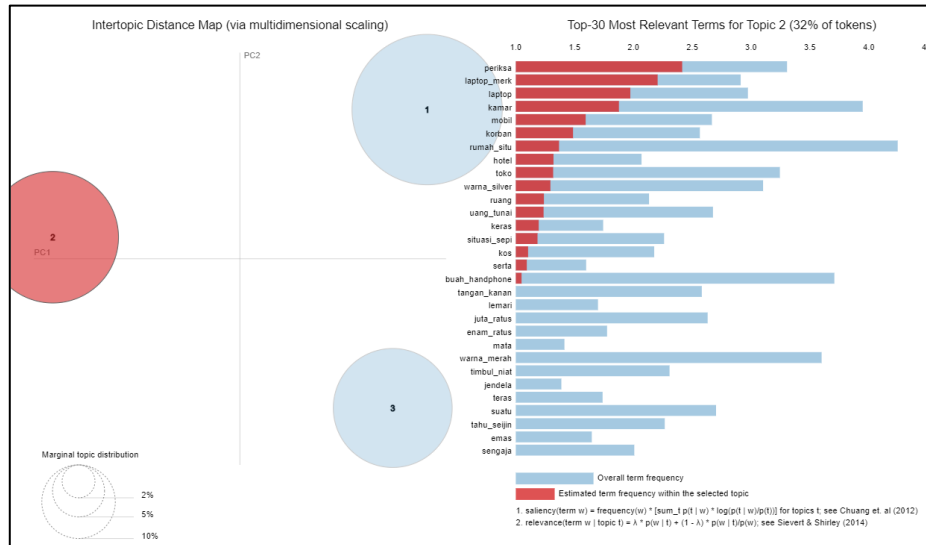
**Gambar 5.12** *Word Cloud* Perkara Pencurian

Pada **Gambar 5.12** merupakan visualisasi *word cloud* dari 3 topik. Semakin banyak frekuensi kata pada data *set*, maka akan semakin besar huruf yang akan ditampilkan pada *word cloud*. Sebaliknya, semakin sedikit frekuensi kata pada data *set* maka semakin kecil huruf yang akan ditampilkan pada *word cloud*. Berdasarkan **Gambar 5.12** terlihat bahwa frekuensi kata yang ditampilkan paling besar terdapat pada kata “tahan\_rutan”, “diam\_panggil” dan “milik\_simpan”. Hasil tersebut sesuai dengan visualisasi *pyLDAvis* yang terdapat pada **Gambar 5.11**. Kemudian, melakukan visualisasi *pyLDAvis* untuk topik 1 seperti berikut.



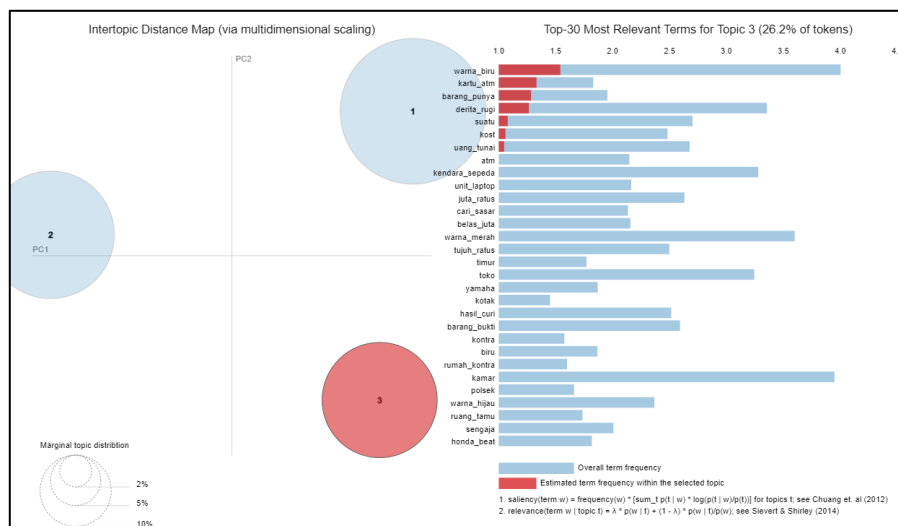
**Gambar 5.13** Visualisasi Topik 1

Pada **Gambar 5.13** visualisasi dari topik 1 ditunjukkan dengan lingkaran berwarna merah dan estimasi *term frequency* topik 1 ditunjukkan dengan diagram batang berwarna merah. *Marginal topic distribution* pada topik 1 mempunyai nilai sebesar 41,8% yang artinya tingkat pentingnya topik 1 terhadap dokumen memuat 41,8% dari keseluruhan topik dalam dokumen. Pada diagram batang berwarna merah memperlihatkan 30 frekuensi kata yang relevan dengan topik 1. Dapat dilihat bahwa 3 kata yang sering muncul pada topik 1 adalah “rumah\_situ”, “buah\_handphone” dan “handphone”.



Gambar 5.14 Visualisasi Topik 2

Pada Gambar 5.14 visualisasi dari topik 2 ditunjukkan dengan lingkaran berwarna merah dan estimasi *term frequency* topik 2 ditunjukkan dengan diagram batang berwarna merah. *Marginal topic distribution* pada topik 2 mempunyai nilai sebesar 32% yang artinya tingkat pentingnya topik 2 terhadap dokumen memuat 32% dari keseluruhan topik dalam dokumen. Pada diagram batang berwarna merah memperlihatkan 17 frekuensi kata yang relevan dengan topik 2. Dapat dilihat bahwa 3 kata yang sering muncul pada topik 2 adalah “periksa”, “laptop\_merk” dan “laptop”.



Gambar 5.15 Visualisasi Topik 3

Pada Gambar 5.15 visualisasi dari topik 3 ditunjukkan dengan lingkaran berwarna merah dan estimasi *term frequency* topik 3 ditunjukkan dengan diagram batang berwarna merah. *Marginal topic distribution* pada topik 3 mempunyai nilai

sebesar 26,2% yang artinya tingkat pentingnya topik 3 terhadap dokumen memuat 26,2% dari keseluruhan topik dalam dokumen. Pada diagram batang berwarna merah memperlihatkan 7 frekuensi kata yang relevan dengan topik 3. Dapat dilihat bahwa 3 kata yang sering muncul pada topik 3 adalah “warna\_biru”, “kartu\_atm” dan “barang\_punya”.

Setelah melakukan visualisasi 3 topik menggunakan *pyLDAvis*, maka dapat disimpulkan kasus tindak pidana yang sering terjadi dan dibahas di Pengadilan Negeri Yogyakarta berdasarkan data dokumen surat dakwaan pada tahun 2019-2021 dengan membuat pemodelan topik. Pemodelan topik dilakukan menggunakan kata-kata atau *term frequency* yang sering muncul dengan cara menyambungkan kata yang relevan dan berkaitan satu sama lain. Berikut adalah pemodelan 3 topik dari data klasifikasi perkara pencurian di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.

**Tabel 5.17** Hasil Pemodelan 3 Topik Pencurian

Topik	Pemodelan Topik
1	0,010*“rumah_situ” + 0,010*“buah_handphone” + 0,009*“handphone” + 0,008*“unit_handphone” + 0,008*“warna_merah” + 0,008*“warna_biru” + 0,007*“pakai_perintah” + 0,007*“ambil_rusak” + 0,007*“kendara_sepeda” + 0,007*“potong_panjat”
2	0,013*“periksa” + 0,012*“laptop_merk” + 0,011*“laptop” + 0,010*“kamar” + 0,009*“mobil” + 0,008*“korban” + 0,008*“rumah_situ” + 0,007*“hotel” + 0,007*“toko” + 0,007*“warna_silver”
3	0,010*“warna_biru” + 0,009*“kartu_atm” + 0,009*“barang_punya” + 0,009*“derita_rugi” + 0,007*“suatu” + 0,007*“kost” + 0,007*“uang_tunai” + 0,006*“atm” + 0,006*“kendara_sepeda” + 0,006*“unit_laptop”

Berdasarkan **Tabel 5.17** hasil pemodelan 3 topik menggunakan metode *Latent Dirichlet Allocation (LDA)* didapatkan kesimpulan bahwa pada klasifikasi perkara pencurian banyak membahas tentang jenis barang yang dicuri, alat pendukung dalam pencurian dan tempat terjadinya peristiwa. Berikut adalah kesimpulan dari masing-masing topik pada klasifikasi perkara pencurian.

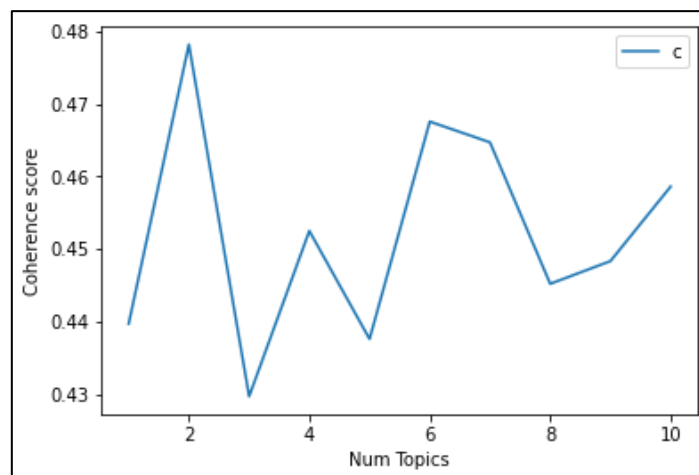
**Tabel 5.18** Kesimpulan Pemodelan 3 Topik Pencurian

Topik	Jenis Barang	Alat Pendukung	Tempat Terjadi
1	<i>Handphone</i>	Mobil	Rumah
2	Laptop	Sepeda Motor	Rumah
3	Uang (Kartu Atm)	Sepeda Motor	Kontrakan

Pada **Tabel 5.18** kesimpulan dari masing-masing pemodelan 3 topik klasifikasi perkara pencurian, dapat diketahui topik 1 banyak membahas tentang barang yang dicuri yaitu *handphone*, dengan alat pendukung dalam pencurian yaitu mobil dan tempat terjadinya peristiwa di rumah. Topik 2 banyak membahas tentang barang yang dicuri yaitu laptop, dengan alat pendukung pencurian yaitu sepeda motor dan tempat terjadinya peristiwa di rumah. Topik 3 banyak membahas tentang barang yang dicuri yaitu uang (kartu atm), dengan alat pendukung pencurian yaitu sepeda motor dan tempat terjadinya peristiwa di kontrakan.

### 5.4.3 Kesehatan

Pada tahap *topic modelling* terlebih dahulu ditentukan jumlah topik menggunakan *coherence score*. *Coherence score* merupakan ukuran yang digunakan sebagai evaluasi dari *topic modelling* yang divisualisasikan dalam bentuk grafik. Semakin besar *coherence score*, maka model yang dihasilkan akan semakin baik. Berikut adalah grafik *coherence score* dengan menggunakan limit topik yaitu 1-11 dari data klasifikasi perkara kesehatan di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.



**Gambar 5.16** Grafik *Coherence Score* Kesehatan



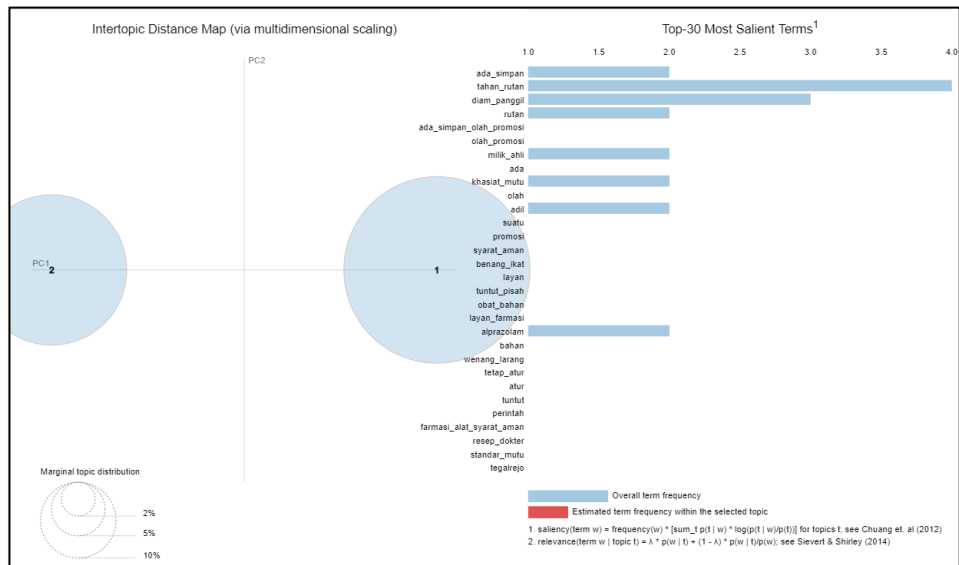
Pada **Gambar 5.16** merupakan visualisasi dari grafik *coherence score* untuk klasifikasi perkara kesehatan, dengan nilai tertinggi terdapat pada topik ke-2. Sehingga jumlah topik yang terbentuk dari klasifikasi perkara kesehatan di Pengadilan Negeri Yogyakarta pada tahun 2019-2021 adalah sebanyak 2. Besarnya *coherence score* atau nilai koherensi dari topik berdasarkan grafik pada **Gambar 5.16** dapat dilihat pada **Tabel 5.19** berikut.

**Tabel 5.19** Nilai Koherensi Topik Kesehatan

<i>Num Topics</i>	<i>Coherence Value</i>	<i>Num Topics</i>	<i>Coherence Value</i>
1	0,4397	6	0,4675
2	0,4782	7	0,4647
3	0,4297	8	0,4452
4	0,4525	9	0,4483
5	0,4376	10	0,4586

Berdasarkan nilai pada **Tabel 5.19** dapat dilihat bahwa *num topics* 2 merupakan *coherence value* tertinggi dengan nilai sebesar 0,4782. Hal ini menunjukkan bahwa jumlah topik yang terbentuk terkait klasifikasi perkara kesehatan untuk mengetahui kasus tindak pidana yang sering terjadi di Pengadilan Negeri Yogyakarta adalah sebanyak 2 topik. Hasil tersebut sama dan sesuai dengan visualisasi grafik *coherence score* pada **Gambar 5.16**. Sehingga jumlah 2 topik yang dihasilkan tersebut akan dijadikan acuan dalam pembuatan model.

Selanjutnya, peneliti melakukan analisis *topic modelling* menggunakan visualisasi dari *pyLDAvis* untuk mengetahui keterkaitan antar kata yang sering muncul pada suatu topik. Berikut adalah visualisasi dari 2 topik menggunakan *pyLDAvis*.



**Gambar 5.17** Visualisasi 2 Topik Perkara Kesehatan

Pada visualisasi *pyLDAvis* menunjukkan diagram terbagi menjadi 2 panel, yaitu panel kiri dan panel kanan. Pada panel kiri merupakan pemetaan jarak antar topik (*intertopic distance map*), divisualisasikan dengan lingkaran-lingkaran berisikan angka yang menunjukkan *num topics*. Besar kecilnya suatu lingkaran menggambarkan seberapa penting dan sering topik tersebut muncul dalam dokumen surat dakwaan. Selain itu dapat dilihat pada visualisasi *intertopic distance map* terbagi menjadi 2 sumbu yaitu *Principal Component 1 (PC1)* dan *Principal Component 2 (PC2)*, sehingga terbentuk kuadran yang memiliki makna bahwa lingkaran topik yang berdekatan memiliki hubungan yang kuat dan lingkaran topik yang berjauhan memiliki hubungan yang lemah. Berikut adalah nilai *principal component* dari 2 topik yang digunakan.

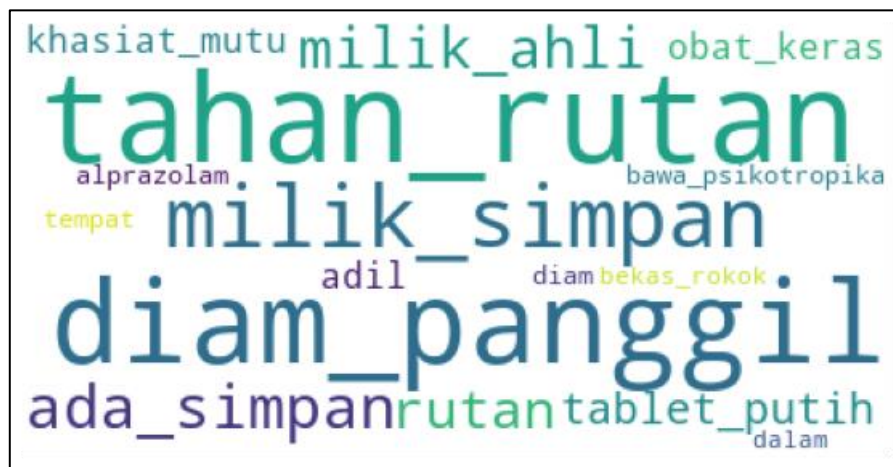
**Tabel 5.20** Hasil *Principal Component*

<i>Num Topics</i>	<b>PC1</b>	<b>PC2</b>
1	0,0080	0,0000
2	-0,0080	0,0000

Berdasarkan **Tabel 5. 12** hasil *Principal Component* dapat dilihat bahwa nilai *PC1* merupakan titik koordinat pada sumbu *X* dan nilai *PC2* merupakan titik kordinat pada sumbu *Y*. Sehingga dengan melihat nilai *PC1* dan *PC2* dapat diketahui bahwa topik ke-1 berada di kuadran I ditandai dengan sumbu *X* bernilai positif dan sumbu *Y* bernilai positif mempunyai arti bahwa representasi kata pada model tinggi,

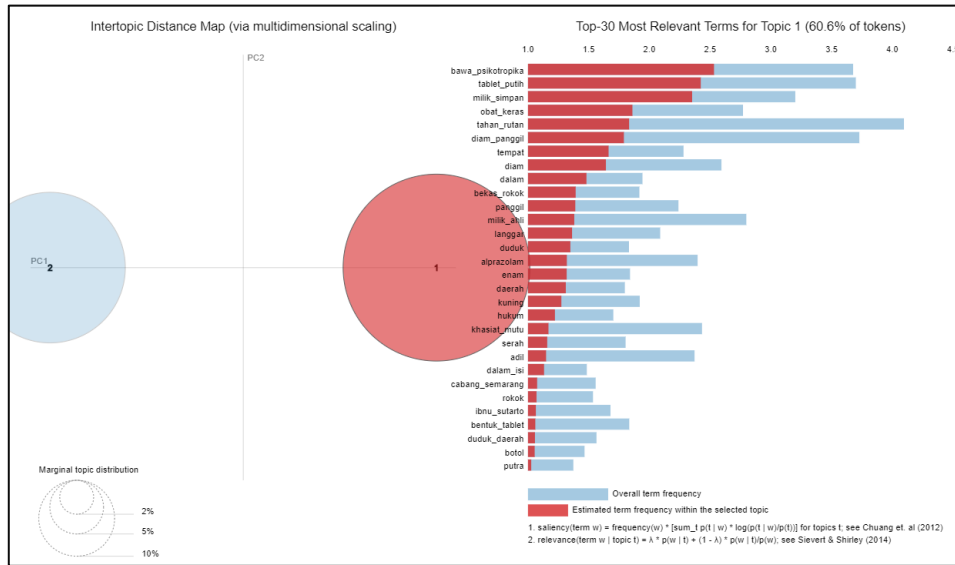
dan mempunyai variasi kata yang tinggi, sehingga berkaitan dengan topik utama. Topik ke-2 berada di kuadran II ditandai dengan sumbu  $X$  bernilai negatif dan sumbu  $Y$  bernilai positif mempunyai arti bahwa representasi kata pada model rendah dan mempunyai variasi kata yang tinggi, namun lebih berkaitan dengan topik tambahan.

Berikut merupakan visualisasi berupa *word cloud* dari 2 topik untuk menampilkan frekuensi kata-kata yang sering muncul pada teks atau dokumen agar mudah dipahami oleh pembaca dari data surat dakwaan untuk klasifikasi perkara kesehatan di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.



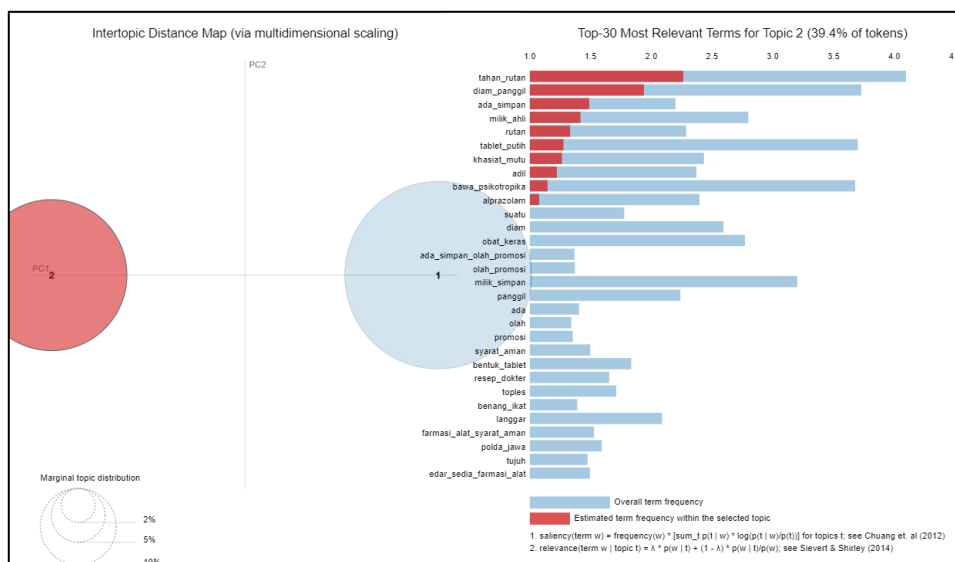
**Gambar 5.18** *Word Cloud* Perkara Kesehatan

Pada **Gambar 5.18** merupakan visualisasi *word cloud* dari 2 topik. Semakin banyak frekuensi kata pada data *set*, maka akan semakin besar huruf yang akan ditampilkan pada *word cloud*. Sebaliknya, semakin sedikit frekuensi kata pada data *set* maka semakin kecil huruf yang akan ditampilkan pada *word cloud*. Berdasarkan **Gambar 5.18** terlihat bahwa frekuensi kata yang ditampilkan paling besar terdapat pada kata “tahan\_rutan”, “diam\_panggil” dan “milik\_simpan”. Hasil tersebut sesuai dengan visualisasi *pyLDAvis* yang terdapat pada **Gambar 5.17**. Kemudian, melakukan visualisasi *pyLDAvis* untuk topik 1 seperti berikut.



**Gambar 5.19** Visualisasi Topik 1

Pada **Gambar 5.19** visualisasi dari topik 1 ditunjukkan dengan lingkaran berwarna merah dan estimasi *term frequency* topik 1 ditunjukkan dengan diagram batang berwarna merah. *Marginal topic distribution* pada topik 1 mempunyai nilai sebesar 60,6% yang artinya tingkat pentingnya topik 1 terhadap dokumen memuat 60,6% dari keseluruhan topik dalam dokumen. Pada diagram batang berwarna merah memperlihatkan 30 frekuensi kata yang relevan dengan topik 1. Dapat dilihat bahwa 3 kata yang sering muncul pada topik 1 adalah “bawa\_psikotropika”, “tablet\_putih” dan “milik\_simpan”.



**Gambar 5.20** Visualisasi Topik 2

Pada **Gambar 5.20** visualisasi dari topik 2 ditunjukkan dengan lingkaran berwarna merah dan estimasi *term frequency* topik 2 ditunjukkan dengan diagram

batang berwarna merah. *Marginal topic distribution* pada topik 2 mempunyai nilai sebesar 39,4% yang artinya tingkat pentingnya topik 2 terhadap dokumen memuat 39,4% dari keseluruhan topik dalam dokumen. Pada diagram batang berwarna merah memperlihatkan 10 frekuensi kata yang relevan dengan topik 2. Dapat dilihat bahwa 3 kata yang sering muncul pada topik 2 adalah “tahan\_rutan”, “diam\_panggil” dan “ada\_simpan”.

Setelah melakukan visualisasi 2 topik menggunakan *pyLDAvis*, maka dapat disimpulkan kasus tindak pidana yang sering terjadi dan dibahas di Pengadilan Negeri Yogyakarta berdasarkan data dokumen surat dakwaan pada tahun 2019-2021 dengan membuat pemodelan topik. Pemodelan topik dilakukan menggunakan kata-kata atau *term frequency* yang sering muncul dengan cara menyambungkan kata yang relevan dan berkaitan satu sama lain. Berikut adalah pemodelan 2 topik dari data klasifikasi perkara kesehatan di Pengadilan Negeri Yogyakarta pada tahun 2019-2021.

**Tabel 5.21** Hasil Pemodelan 2 Topik Kesehatan

Topik	Pemodelan Topik
1	0,014*“bawa_psikotropika” + 0,014*“tablet_putih” + 0,013*“milik_simpan” + 0,010*“obat_keras” + 0,010*“tahan_rutan” + 0,010*“diam_panggil” + 0,009*“tempat” + 0,009*“diam” + 0,008*“dalam” + 0,008*“bekas_rokok”
2	0,019*“tahan_rutan” + 0,017*“diam_panggil” + 0,013*“ada_simpan” + 0,012*“milik_ahli” + 0,011*“rutan” + 0,011*“tablet_putih” + 0,011*“khasiat_mutu” + 0,011*“adil” + 0,010*“bawa_psikotropika” + 0,009*“alprazolam”

Berdasarkan **Tabel 5.21** hasil pemodelan 2 topik menggunakan metode *Latent Dirichlet Allocation (LDA)* didapatkan kesimpulan bahwa pada klasifikasi perkara kesehatan banyak membahas tentang jenis kasus yang berhubungan dengan ahli kesehatan, dan jenis obat yang dimiliki tanpa izin. Berikut adalah kesimpulan dari masing-masing topik pada klasifikasi perkara pencurian.

**Tabel 5.22** Kesimpulan Pemodelan 2 Topik Kesehatan

Topik	Jenis Kasus	Jenis Obat
1	Menyimpan	Obat Keras
2	Menyimpan	Obat Milik Ahli(Tanpa Izin)

Pada **Tabel 5.22** kesimpulan dari masing-masing pemodelan 2 topik klasifikasi perkara kesehatan, dapat diketahui topik 1 banyak membahas tentang kasus menyimpan obat keras yang dilarang oleh ahli kesehatan. Topik 2 banyak membahas tentang kasus menyimpan obat milik ahli kesehatan tanpa izin.

Berdasarkan penelitian diatas dapat disimpulkan bahwa sebanyak 922 data kasus tindak pidana dipublikasikan dan dapat diakses lihat oleh masyarakat luas, dengan kasus tertinggi banyak terjadi pada tahun 2021. Terdapat 3 klasifikasi perkara kasus tindak pidana tertinggi yang sering terjadi yaitu narkoba, pencurian dan kesehatan. Dari 3 klasifikasi perkara kasus tindak pidana tersebut banyak mendapatkan hukuman penjara waktu kurang dari satu tahun dengan tidak mendapatkan denda, hal itu dikarenakan perlu mempertimbangkan berbagai faktor seperti kondisi ekonomi, berat dan ringannya kasus, serta kerugian yang ditimbulkan.

Pada hasil analisis *topic modelling* berdasarkan 3 klasifikasi perkara kasus tindak pidana tertinggi untuk mengetahui topik yang sering terjadi dan dibahas didapatkan untuk klasifikasi perkara narkoba sebanyak 3 topik, klasifikasi perkara pencurian sebanyak 3 topik dan klasifikasi perkara kesehatan sebanyak 2 topik.

## **BAB VI**

### **PENUTUP**

#### **6.1. Kesimpulan**

Berdasarkan hasil analisis yang telah dilakukan peneliti sebelumnya, maka diperoleh beberapa kesimpulan sebagai berikut :

1. Didapatkan gambaran umum dari data kasus tindak pidana di Pengadilan Negeri Yogyakarta seperti berikut :
  - a. Gambaran umum jumlah publikasi kasus tindak pidana di Pengadilan Negeri Yogyakarta pada tahun 2019-2021 yaitu sebanyak 971 kasus tindak pidana, dengan 922 kasus dapat diakses oleh masyarakat luas dan 49 kasus tidak dapat diakses oleh masyarakat luas. Serta, jumlah kasus tindak pidana tertinggi terdapat pada tahun 2021 sebanyak 326 kasus. Sedangkan di tahun 2019 sebanyak 325 kasus dan tahun 2020 sebanyak 320 kasus.
  - b. Gambaran umum 3 klasifikasi perkara tertinggi dari data kasus tindak pidana di Pengadilan Negeri Yogyakarta pada tahun 2019-2021 yaitu Narkotika, Pencurian dan Kesehatan.
  - c. Gambaran umum hukuman dan denda yang didapatkan pada kasus tindak pidana di Pengadilan Negeri Yogyakarta pada tahun 2019-2021 tertinggi terdapat pada hukuman penjara waktu kurang dari satu tahun dengan tidak mendapatkan denda di klasifikasi perkara narkotika yaitu sebanyak 450 kasus.
2. Didapatkan hasil analisis *topic modelling* menggunakan metode *Latent Dirichlet Allocation (LDA)* untuk mengetahui topik yang sering terjadi dan dibahas pada data kasus tindak pidana di Pengadilan Negeri Yogyakarta pada tahun 2019-2021 berdasarkan 3 klasifikasi perkara tertinggi yang terdiri dari narkotika, pencurian dan kesehatan adalah sebagai berikut :
  - a. Klasifikasi perkara narkotika terdapat 3 topik yang sering terjadi dan dibahas yaitu jenis kasus yang terjadi, zat narkotika yang digunakan dan tempat terjadinya peristiwa.

- b. Klasifikasi perkara pencurian terdapat 3 topik yang sering terjadi dan dibahas yaitu jenis barang yang dicuri, alat pendukung pencurian dan tempat terjadinya peristiwa.
- c. Klasifikasi perkara kesehatan terdapat 2 topik yang sering terjadi dan dibahas yaitu jenis kasus yang berhubungan dengan ahli kesehatan dan jenis obat yang dimiliki tanpa izin.

## 6.2. Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat saran yang dapat penulis sampaikan yaitu :

1. Pada penelitian ini, peneliti menggunakan data kasus tindak pidana pada tahun 2019-2021. Untuk penelitian selanjutnya dapat menambahkan dan menggunakan data terbaru, sehingga mendapatkan hasil yang lebih baik
2. Pada penelitian ini, peneliti menggunakan metode *Latent Dirichlet Allocation (LDA)*. Untuk penelitian selanjutnya dapat dikembangkan analisis dengan penambahan metode topik.
3. Pada penelitian selanjutnya perlu diperhatikan pengoptimalkan parameter model *Latent Dirichlet Allocation (LDA)* untuk mendapatkan hasil representasi topik yang lebih baik.
4. Untuk aparat penegak hukum di daerah Yogyakarta perlu ditingkatkan lagi upaya dalam menangani dan menekan tingginya kasus tindak pidana dengan memberikan hukuman yang tegas agar para pelaku tidak mengulangi perbuatannya. Khususnya perkara narkoba yang banyak terjadi di rutan, perkara pencurian yang banyak terjadi di rumah dan perkara kesehatan yang memiliki obat tanpa izin.
5. Untuk masyarakat umum perlu ditingkatkan lagi kewaspadaan dan pemahaman dalam menghindari resiko menjadi korban kasus tindak pidana.



## DAFTAR PUSTAKA

- Agatha Deolika, K. K. (2019 ). Analisis Pembobotan Kata Pada Klasifikasi Text Mining. *Jurnal Teknologi Informasi* .
- Ahmad, A. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network dan Deep Learning. *Teknologi Indonesia*, 4-5.
- Akhsin Nurlayli, M. A. (2019). Topic Modeling Penelitian Dosen JPTEI UNY pada Google Scholar Menggunakan Latent Dirichlet Allocation. *ELIVO*, 154-161.
- Blei, D. M., Ng, A. Y., & Jordan, M. (2003). Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data. *Machine Learning Research*, 3, 993–1022. <https://doi.org/10.1016/B978-0-12-411519-4.00006-9>
- Blei, D., Carin, L., & Dunson, D. (2010). Probabilistic Topic Models. *IEEE Signal Process Magazine*, 27(6), 55–65.
- Cendana, M., & Permana, D. H. (2019). Pra-Pemrosesan Teks Pada Grup Whatsapp Untuk Pemodelan Topik. *Mantik Penusa*, 3, 112.
- Chazawi, A. (2008). Pelajaran Hukum Pidana 1. Jakarta: PT. Raja Grafindo.
- Debby Alita, A. R. (2020). Pendeteksian Sarkasme pada Proses Analisis Sentimen Menggunakan Random Forest Classifier. *Jurnal Komputas*, 50-58.
- Gata, W. (2017). Akurasi Text Mining Menggunakan Algoritma K-Nearest Neighbour pada Data Content Berita SMS. 6, 1-13.
- Gunarto, M. P. (2009). Sikap Memidana yang Berorientasi pada Tujuan Pidanaan. *Mimbar Hukum*, 21,108.
- Harjanta, A. T. (2015). Preprocessing Text untuk Meminimalisir Kata yang Tidak Berarti dalam Proses Text Mining. *Journal Informatika Upgris Universitas PGRI* , 1-9.
- Hikmah, N. (2018 ). Pemanfaatan Text Mining dalam Pencarian Ayat Al-Quran menggunakan TF-IDF dan Cosine Similarity. *Jurnal Antartika* , 13-22.
- Indroharto. (2003). Usaha Memahami Undang-Undang Tentang Peradilan Tata Usaha Negara Buku II beracara di Pengadilan Tata Usaha Negara. Jakarta: Pustaka Sinar Harapan.
- Kannitha, D. Z. T., Mustafid, M., & Kartikasari, P. (2022). Pemodelan Topik Pada

- Keluhan Pelanggan Menggunakan Algoritma Latent Dirichlet Allocation Dalam Media Sosial Twitter. *Jurnal Gaussian*, 11(2), 266–277. <https://doi.org/10.14710/j.gauss.v11i2.35474>
- Lamintang, P. (1984). Hukum Penitensier Indonesia. Bandung: Armico.
- Leon, Y. (2021, 12 29). *Kasus Kriminal Ada Kenaikkan Sepanjang 2021*. Diambil kembali dari Solopos.com: <https://www.solopos.com/polda-diy-kasus-kriminal-ada-kenaikkan-sepanjang-2021-1226348>
- Marpaung, L. (1992). Proses Penanganan Perkara Pidana Bagian Pertama Penyelidikan dan Penyidikan. Jakarta: Sinar Grafika.
- Mertokusuma, S. (1986). Norma dan Kaidah Hukum. Yogyakarta: Offset Gajdah Mada University Press.
- Mutsani, H. (2019, Desember 01). *Statistika Deskriptif : Pengertian, Karakteristik, Contoh, Fungsi, Dan Ruang Lingkupnya Lengkap*. Diambil kembali dari tutorialbahasainggris:<https://tutorialbahasainggris.co.id/statistika-deskriptif-pengertian-karakteristik-contoh-fungsi-dan-ruang-lingkupnya-lengkap/>
- Nasution, A. K. (1972). Masalah Surat Tuduhan dalam Proses Pidana. Jakarta: Percetakan Negara.
- Nugraha, J. (2015). Advances on Statistical Methods in Engineering, Science, Economy, . *Proceedings of The First International Conference on Statistical Methods in Engineering, Science, Economy and Education* , 66-71.
- Pratama, T. R., Richasdy, D., & Purbolakson, M. D. (2022). Telkom University Opinion Topic Modeling on Twitter Using Latent Dirichlet Allocation During Covid-19 Pandemic. *Media Informatika Budidarma*, 6, 1816–1825. <https://doi.org/10.30865/mib.v6i4.4426>
- Putra, I. m., & Kusumawardani, R. P. (2017). Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA). *Teknik ITS*, 6, 311-316.
- Rakhmawati, N. A., Waskitho, R. B., Rahman, D. A., & Nuha, M. F. A. U. (2021). Klasterisasi Topik Konten Channel Youtube Gaming Indonesia Menggunakan Latent Dirichlet Allocation. *Journal of Information Engineering and Educational Technology*, 5(2), 78–83. <https://doi.org/10.26740/jieet.v5n2.p78-83>

- Sari, I. M., Efendi, R., & Chairani, L. (2020). Hubungan Sistem Pembelajaran Daring Di Era COVID-19 Terhadap Kesehatan Mental Guru SD : Uji Chi-Square dan Dependency Degree. *Seminar Nasional Teknologi Informasi, Komunikasi Dan Industri (SNTIKI) 12*, 624–631.
- Sahria, Y., & Hatta Fudholi, D. (2020). Analisis Topik Penelitian Kesehatan di Indonesia Menggunakan Metode Topic Modeling LDA (Latent Dirichlet Allocation). *Rekayasa Sistem Dan Teknologi*, 1(3), 336–344.
- Sastriana, Mahyulis, H., Nurmayanti, W. P., & Hakim, L. (2021). Implementation of Topic Modeling with Latent Dirichlet Allocation (LDA) on the Judge's Decision. *Isst*, 5(October), 113–128.
- Scholtes, J. C. (2009). Text Mining : The next step in Search Technology.
- Septiani, N., Erwansyah, K., Kom, S., Kom, M., Suryanata, M. G., Kom, S., & Kom, M. (2020). *ANALISIS DATA MINING PENGELOMPOKKAN KASUS MEDAN POLONIA DENGAN MENGGUNAKAN METODE K-MEANS CLUSTERING*. 3(2), 1–13.
- Setyawati, D. (2011). Metode Analisis Crosstabs (Chi-Square) Untuk Meneliti Hubungan Antara Pekerjaan Dengan Jenis Kasus Pidana Dan Hubungan Antara Wilayah Dengan Jenis Kasus Pidana Di Kabupaten Semarang (Studi Kasus di Kejaksaan Negeri Ambarawa). Yogyakarta: Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Islam Indonesia.
- Setijohatmo, U. T., Rachmat, S., Susilawati, T., Rahman, Y., & Kunci, K. (2020). Analisis Metoda Latent Dirichlet Allocation untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik. *Prosiding The 11th Industrial Research Workshop and National Seminar*, 402–408.
- Sholehuddin. (2003). Sistem Sanksi Dalam Hukum Pidana : Ide Dasae Double Track System dan Implementasinya. Jakarta: Raja Grafindo Persada.
- Sinaga, A. S. (2020). Machine Learning Prediksi Karakter Pengguna Hastag (#) Bahasa Generasi Milenial Di Sosial Media. *Indonesian Journal of Applied*, 166-167.
- Sujarweni, W. (2012). Statistika Untuk Penelitian. Yogyakarta: Graha Ilmu.

- Suwanto Sanjaya, E. A. (2015). Pengelompokan Dokumen Menggunakan Winnowing Fingerprint dengan Metode K-Nearest Neighbour. *Jurnal Hasil Penelitian Ilmu Komputer dan Teknologi Informasi*, 50-56.
- Timbuleng, A. E. (2020). Tindak Pidana Di Bidang Perizinan Menurut Undang-Undang Nomor 11 Tahun 2010 Tentang Cagar Budaya. *Lex Crimen*, 28-36.
- Tresnasari, N. A., Adji, T. B., & Permanasari, A. E. (2020). Social-Child-Case Document Clustering based on Topic Modeling using Latent Dirichlet Allocation. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 14(2), 179. <https://doi.org/10.22146/ijccs.54507>
- Wahyuni, F. (2017). Dasar-dasar Hukum Pidana Di Indonesia. Tangerang Selatan: PT Nusantara Persada Utama.
- Weda, M. D. (1996). Kriminologi. Jakarta: PT. Raja Grafindo Persada.
- Widi, S. (2022, Oktober 28). *Polri Catat 274.988 Kasus Tindak Pidana pada 2021*. Diambil kembali dari DataIndonesia.id: <https://dataindonesia.id/ragam/detail/polri-catat-274988-kasus-tindak-pidana-pada-2021>
- Willy, A., & Hartono, J. (2016). Partial Least Square (PLS) Alternatif Structural Equation Modeling (SEM) dalam Penelitian Bisnis. Yogyakarta: Andi.
- Yogyakarta, P. N. (2019). *Tugas Pokok dan Fungsi*. <https://www.pn-yogyakarta.go.id/pnyk/tentang-pengadilan/2015-06-22-15-58-25.html>
- Yoren. (2018). Perbandingan Raw TF Dan Binary TF Pada Sistem Pencarian Di Situs Museum Wayang The Comparison Of Raw TF And Binary TF On Searching System In Website Of Kekayon Museum Of Puppets In Yogyakarta.

## LAMPIRAN

**Lampiran 1** Data Kasus Tindak Pidana

[https://bit.ly/LampiranData\\_TA\\_DelaGustiara](https://bit.ly/LampiranData_TA_DelaGustiara)