

**IMPLEMENTASI DETEKSI TOPIK PUTUSAN HAKIM
DENGAN *LATENT DIRICHLET ALLOCATION* (LDA)**

(Studi Terhadap Putusan Tindak Pidana di Pengadilan Negeri Sleman, D.I.
Yogyakarta)

TUGAS AKHIR

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



Disusun Oleh:

Fauziah Dewi Nisrina

16611127

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

2020

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul : IMPLEMENTASI DETEKSI TOPIK PUTUSAN
HAKIM DENGAN LATENT DIRICHLET
ALLOCATION (LDA)

Nama Mahasiswa : Fauziah Dewi Nisrina

NIM : 16611127

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

Yogyakarta, April 2020

Pembimbing



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

HALAMAN PENGESAHAN

TUGAS AKHIR

IMPLEMENTASI DETEKSI TOPIK PUTUSAN HAKIM DENGAN *LATENT DIRICHLET ALLOCATION (LDA)*

(Studi Terhadap Putusan Tindak Pidana di Pengadilan Negeri Sleman, D.I.
Yogyakarta)

Nama Mahasiswa : Fauziah Dewi Nisrina

NIM : 16611127

TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL: 12 Mei 2020

Nama Penguji:

Tanda Tangan

1. Karimatul Ummah, S.H., M.Hum.
2. Dina Tri Utari, S.Si., M.Sc.
3. Ayundyah Kesumawati, S.Si., M.Si.

Mengetahui,
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia



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

KATA PENGANTAR



Assalamu'alaikum Wr.Wb

Alhamdulillah puji syukur atas kehadiran Allah SWT yang telah melimpahkan rahmat, hidayah, dan inayah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “Implementasi Deteksi Topik Putusan Hakim Dengan *Latent Dirichlet Allocation* (LDA)” ini dengan lancar dan sebaik-baiknya.

Dalam penyusunan Tugas Akhir ini, penulis menyadari bahwa penulisan laporan ini banyak memperoleh bantuan dari berbagai pihak, baik yang berupa saran, kritik, bimbingan maupun bantuan lainnya. Oleh karena itu pada kesempatan kali ini penulis menyampaikan ucapan 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 Statistikabeserta seluruh jajarannya.
3. Ibu Ayundyah Kesumawati, S.Si., M.Si., selaku dosen pembimbing yang telah meluangkan waktu untuk membimbing penulis selama penyusunan Tugas Akhir ini.
4. Seluruh staff pengajar Program Studi Statistika yang telah memberikan bekal ilmu kepada penulis.
5. Orang tua tercinta beserta keluarga besar yang telah memberikan doa, dukungan, dan motivasi kepada penulis.
6. Rizky Desi, Alfi Indah, Rizqi Ananda, dan Widia Nusalina yang telah membantu dan menyuport hingga selesai penulisan Tugas Akhir ini.
7. Teman-teman seperjuangan sekaligus satu bimbingan Tugas Akhir yaitu cinmey, dea, ella, kides, laras, dan mita yang selalu berbagi ilmu dan berbagi cerita serta pengalaman.
8. Semua teman-teman statistika angkatan 2016 yang selalu menjadi semangat bagi penulis.

9. Serta semua pihak yang tidak dapat disebutkan satu-persatu telah membantu penulis menyelesaikan laporan kerja praktik ini baik secara langsung maupun tidak langsung.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari kata sempurna. Oleh karena itu penulis mengharapkan kritik dan saran yang membangun dari pembaca untuk menyempurnakan penulisan Tugas Akhir ini. Penulis berharap semoga penulisan Tugas Akhir ini dapat bermanfaat bagi semua pihak yang membutuhkan. Demikianlah yang dapat disampaikan, semoga Allah SWT senantiasa melimpahkan rahmat dan ridho-Nya kepada semua pihak yang telah membantu penulis.

Wassalamualaikum Wr.Wb

Yogyakarta, Maret 2020

(Fauziah Dewi Nistrina)



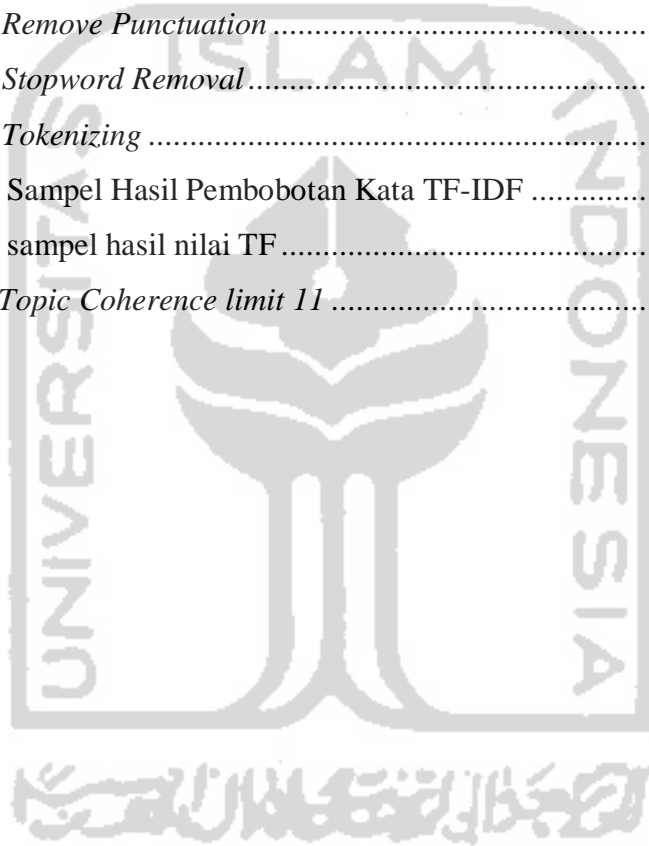
DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN PERSETUJUAN PEMBIMBING.....	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR	iv
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN.....	x
PERNYATAAN	xi
INTISASI	xii
ABSTRACT.....	xiii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	4
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
BAB 2 TINJAUAN PUSTAKA.....	6
BAB 3 LANDASAN TEORI.....	10
3.1 Kejahatan.....	10
3.2 Tindak Pidana	10
3.2.1 Pertanggungjawaban Pidana	11
3.2.2 Jenis-jenis Pidana	11
3.2.3 Acara Pemeriksaan	13
3.3 Hukum Pidana	14
3.4 Statistika Deskriptif	14
3.5 <i>Artificial Intelligence</i>	14
3.6 <i>Machine Learning</i>	16
3.7 <i>Text Mining</i>	16

3.12	<i>Text Preprocessing</i>	17
3.13	Pembobotan Kata (<i>Term Weighting</i>)	17
3.14	<i>Topic Modeling</i>	19
3.15	<i>Latent Dirichlet Allocation (LDA)</i>	20
3.16	<i>Topic Coherence</i>	22
BAB 4	METODOLOGI PENELITIAN	23
4.1	Populasi dan Sampling Penelitian.....	23
4.2	Jenis dan Sumber Data.....	23
4.3	Variabel Penelitian.....	23
4.4	Metode Analisis Data.....	24
4.5	Tahapan Penelitian.....	24
BAB 5	HASIL DAN PEMBAHASAN	26
5.1	Analisis Deskriptif	26
5.2	<i>Preprocessing Data</i>	28
5.3	Pembobotan Kata.....	33
5.4	<i>Topic Modeling</i> dengan <i>Latent Dirichlet Allocation (LDA)</i>	35
BAB 6	PENUTUP.....	46
6.1	Kesimpulan.....	46
6.2	Saran.....	46
	DAFTAR PUSTAKA.....	48
	LAMPIRAN	53

DAFTAR TABEL

Tabel 2.1. Perbandingan dengan Penelitian Terdahulu	8
Tabel 4.1. Definisi Operasional Variabel	23
Tabel 5.1 Data Awal Penelitian	28
Tabel 5.2 Hasil <i>Lower Casing</i>	29
Tabel 5.3 Hasil <i>Remove Punctuation</i>	30
Tabel 5.4 Hasil <i>Stopword Removal</i>	31
Tabel 5.5 Hasil <i>Tokenizing</i>	32
Tabel 5.6 Tabel Sampel Hasil Pembobotan Kata TF-IDF	33
Tabel 5.7 Tabel sampel hasil nilai TF	34
Tabel 5.8 Nilai <i>Topic Coherence limit 11</i>	37



DAFTAR GAMBAR

Gambar 1.1. Selang Waktu Terjadinya Tindak Kejahatan (<i>Crime Clock</i>) di D.I. Yogyakarta, 2016-2018.....	2
Gambar 3.1. Penerapan Konsep Kecerdasan Buatan di Komputer.....	15
Gambar 3.2. Representasi Model LDA	20
Gambar 3.3. Representasi Model LDA	21
Gambar 4.1. <i>Flowchart</i> Tahapan Penelitian	25
Gambar 5.1. Persentase Jenis Kelamin Tindak Pidana	26
Gambar 5.2. Jumlah Perkara Tindak Pidana Berdasarkan Klasifikasi Perkara Setelah Putusan	27
Gambar 5.3. Grafik <i>Coherence Score</i>	36
Gambar 5.4. Visualisasi Topic Modeling dengan <i>PyLDAvis</i>	37
Gambar 5.5. Visualisasi Topik 1.....	38
Gambar 5.6. Model Topik 1.....	39
Gambar 5.7. <i>Word Cloud</i> Topik 1.....	39
Gambar 5.8. Visualisasi Topik 2.....	40
Gambar 5.9. Model Topik 2.....	40
Gambar 5.10. <i>Word Cloud</i> Topik 2.....	41
Gambar 5.11. Visualisasi Topik 3.....	41
Gambar 5.12. Model Topik 3.....	42
Gambar 5.13. <i>Word Cloud</i> Topik 3.....	42
Gambar 5.14. Visualisasi Topik 4.....	43
Gambar 5.15. Model Topik 4.....	43
Gambar 5.16. <i>Word Cloud</i> Topik 4.....	44

DAFTAR LAMPIRAN

Lampiran 1 Data perkara Pengadilan Negeri Sleman tahun 2016 – Januari 2020	54
Lampiran 2 <i>Script dan Output Topic Modeling</i>	55
Lampiran 3 <i>Topic Coherence limit 21, 31, 41, dan 51</i>	63



PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat 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, April 2020



Fauziah Dewi Nisrina

**IMPLEMENTASI DETEKSI TOPIK PUTUSAN HAKIM DENGAN
LATENT DIRICHLET ALLOCATION (LDA)
(Studi Terhadap Putusan Tindak Pidana di Pengadilan Negeri Sleman, D.I.
Yogyakarta)**

Fauziah Dewi Nisrina

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

INTISARI

Kejahatan merupakan salah satu permasalahan yang sering terjadi di kota besar, salah satunya yaitu Daerah Istimewa Yogyakarta. Menurut Badan Pusat Statistik (BPS) D.I. Yogyakarta tahun 2018, Kabupaten Sleman merupakan kabupaten yang memiliki tingkat kriminalitas paling tinggi di Provinsi D.I. Yogyakarta. Salah satu upaya penanggulangan masalah tindak kejahatan, ditetapkan hukum pidana dengan sanksinya berupa pidana yang jenis-jenis pidana tersebut tertuang dalam KUHP Pasal 10. Untuk mengadili individu/kelompok yang melakukan pelanggaran terhadap hukum, diadili melalui pengadilan negeri dimana kejahatan itu dilakukan (*lokus delictie*). Dalam hal ini hakim yang memutuskan sanksi pidana penjara dan/atau denda kepada individu/kelompok tindak pidana dengan hasilnya berupa putusan hakim. Putusan hakim dalam hal ini berupa kalimat kurang dimengerti oleh khalayak umum. Berdasarkan hasil tersebut peneliti ingin membantu khalayak umum/pembaca mengetahui bahasan yang berasal dari putusan hakim tersebut. Salah satu perkembangan teknologi yang berkaitan dengan *text* dapat dimanfaatkan untuk mendapatkan topik dari kumpulan kalimat. Salah satu metode yang dapat digunakan adalah *Latent Dirichlet Allocation* (LDA). Fokus penelitian ini adalah untuk mendapatkan topik dari putusan hakim terkait kejahatan di Pengadilan Negeri Sleman, D.I. Yogyakarta tahun 2016 – Januari 2020. Dari hasil analisis deskriptif jenis kelamin pria yang paling banyak melakukan kejahatan. Sedangkan kejahatan yang paling banyak terjadi adalah pencurian. Kemudian hasil analisis *topic modeling* menggunakan LDA diperoleh 4 topik dengan bahasan pada Topik 1 cenderung membahas kasus perjudian, topik 2 cenderung membahas kasus pencurian/penggelapan/penipuan, topik 3 cenderung membahas kasus narkoba, sedangkan topik 4 cenderung membahas kasus senjata api dan benda tajam / kekerasan/ penganiayaan.

Kata Kunci: Kejahatan, tindak pidana, hukum pidana, *topic modeling*, *Latent Dirichlet Allocation* (LDA).

**IMPLEMENTATION OF TOPICS DETECTION THE JUDGE'S DECISION
WITH LATENT DIRICHLET ALLOCATION (LDA)**

(Study of Criminal Decisions in the Sleman District Court, D.I. Yogyakarta)

Fauziah Dewi Nisrina

Department Statistics, Faculty of Mathematics and Natural Science

Islamic University of Indonesia

ABSTRACT

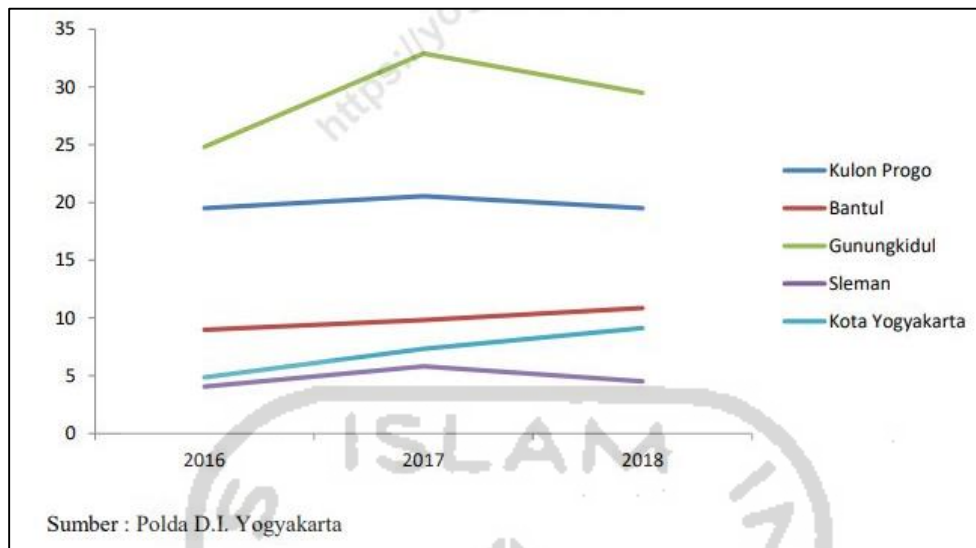
Crime is one of the problems that often occurs in big cities, one of which is Yogyakarta Special Region. According to the Badan Pusat Statistik (BPS) D.I. Yogyakarta in 2018, Sleman Regency is a district that has the highest crime rate in D.I. Yogyakarta. One of the efforts to overcome the problem of crime, a criminal law is established with sanctions in the form of a criminal which the types of crimes are contained in the Criminal Code Article 10. To prosecute individuals/groups who violate the law, tried through the district court where the crime was committed (locus delictie). In this case the judge decides the sanctions imprisonment and/or fines to individuals/groups of criminal acts with the results judge's decision. Judge's decisions in this case are sentences that are not well understood by the general public. Based on these results the researcher wants to help the general public/reader know the discussion that comes from the judge's decision. One of the technological developments related to text can be used to get the topic from a collection of sentences. One of method that can be used is Latent Dirichlet Allocation (LDA). The focus of this research is to get the topic of the judge's decision related to crime in the Sleman District Court, D.I. Yogyakarta in 2016 – January 2020. Form the results of a descriptive analysis of the sex of men who commit the most crimes. While the most common crime is theft. Then the analysis of topic modeling using LDA obtained 4 topics with the discussion on 1st topic tends to discuss gambling cases, 2nd topic tends to discuss theft/embezzlement/fraud cases, 3th topic tends to discuss narcotics cases, while 4th topic tends to discuss cases of firearms and sharp objects/violence/abuse.

Keywords: *Crime, Criminal Act, Criminal Law, Topic Modeling, Latent Dirichlet Allocatin (LDA)*

BAB 1 PENDAHULUAN

1.1 Latar Belakang Masalah

Kejahatan merupakan salah satu permasalahan yang sering terjadi pada kota-kota besar di Indonesia. Menurut Numbeo dalam jurnal (Putri & Susilo, 2019) Indonesia berada di peringkat 68 dari 147 negara berdasarkan peringkat indeks kejahatan pada tahun 2015. Indikator negatif merupakan statistik dan indikator yang biasa digunakan untuk mengukur rasa aman masyarakat, misalnya jumlah angka kejahatan (*crime total*), jumlah orang yang berisiko terkena tindak kejahatan (*crime rate*) setiap 100.000 penduduk. Indikasi masyarakat merasa tidak aman yaitu semakin tinggi angka kriminalitas menunjukkan semakin banyak tindak kejahatan pada masyarakat (Keamanan, 2014). Tindak kejahatan atau kriminalitas di suatu wilayah merupakan suatu perilaku menyimpang yang ada di masyarakat. Perilaku yang menyimpang tersebut merupakan suatu ancaman terhadap norma-norma sosial yang dapat menimbulkan ketegangan individu maupun ketegangan sosial dan merupakan ancaman bagi berlangsungnya ketertiban sosial (Wibowo, 2019). Menurut Kartono D.K. dalam jurnal (Ningsah & Kuncoro, 2017) kriminalitas bukanlah suatu peristiwa bawaan sejak lahir, bukan pula sesuatu yang dapat diwariskan secara biologis. Perilaku kriminal dapat dilakukan oleh siapapun, pria maupun wanita, baik dari usia anak-anak, dewasa, maupun usia lanjut. Tindak kriminal dapat dilakukan secara sadar yaitu dengan dipikirkan, direncanakan, dan diarahkan terlebih dahulu untuk suatu maksud tertentu dalam keadaan yang benar-benar sadar. Selain itu dapat terjadi dalam keadaan setengah sadar yaitu dipengaruhi oleh dorongan dan obsesi yang kuat dalam diri seseorang. Suatu kejahatan dapat pula terjadi secara tidak sadar sama sekali, misalkan karena terpaksa untuk tetap bertahan hidup dengan cara melawan dan membalas menyerang yang pada akhirnya terjadi peristiwa pembunuhan.



Gambar 1.1. Selang Waktu Terjadinya Tindak Kejahatan (*Crime Clock*) di D.I. Yogyakarta, 2016-2018
Sumber: (Wibowo, 2019)

Daerah Istimewa Yogyakarta merupakan salah satu yang memiliki tingkat kejahatan yang tergolong tinggi. Jenis kejahatan yang paling banyak terjadi di D.I. Yogyakarta menurut Badan Pusat Statistik (BPS) Yogyakarta pada tahun 2018 adalah kejahatan pencurian dengan persentase sebanyak 67,12 persen. Gambar 1.1 merupakan gambar selang waktu terjadinya tindak kejahatan (*crime clock*) di D.I. Yogyakarta tahun 2016-2018. Semakin lama selang waktu terjadinya tindak kejahatan, maka semakin sedikit atau berkurangnya kasus tindak kejahatan baru. Sumbu x pada gambar 1.1 menunjukkan tahun terjadinya tindak kejahatan, sedangkan sumbu y menunjukkan jam terjadinya tindak kejahatan. Berdasarkan gambar 1.1 tersebut Kabupaten Sleman merupakan kabupaten yang memiliki tingkat kriminalitas paling tinggi di provinsi D.I. Yogyakarta. Selang waktu terjadinya kejahatan di Kabupaten Sleman menurut Poda D.I. Yogyakarta yaitu 4 jam 31 menit, artinya setiap 4 jam 31 menit terjadi kasus tindak kejahatan baru (Wibowo, 2019).

Dalam rangka menanggulangi tindak kejahatan telah banyak dilakukan berbagai macam cara. Salah satu upaya penanggulangan masalah tindak kejahatan tersebut, telah ditetapkan hukum pidana dengan sanksinya berupa pidana yang

jenis-jenis pidana tersebut tertuang dalam KUHP Pasal 10. Untuk mengadili individu/kelompok yang melakukan kejahatan terhadap hukum diadili melalui pengadilan negeri dimana kejahatan tersebut dilakukan (*lokus delictie*). Begitu pula di daerah Kabupaten Sleman, jika kejahatan tersebut terjadi di daerah Kabupaten Sleman maka akan diadili melalui Pengadilan Negeri Sleman. Tugas dan fungsi Pengadilan Negeri Sleman (PN Sleman) adalah sebagai pelaksana kekuasaan kehakiman pada peradilan tingkat pertama, yang bertugas menyelenggarakan peradilan guna menegakkan hukum dan keadilan berdasarkan Pancasila, dengan tugas pokok menerima, memeriksa, dan mengadili serta menyelesaikan setiap perkara yang diajukan kepada PN Sleman dan tugas lainnya yang diberikan kepada PN Sleman berdasarkan peraturan perundang-undangan (Sleman, 2015). Dalam hal ini hakim yang memutuskan sanksi pidana kepada individu/kelompok tindak pidana dengan hasilnya berupa putusan hakim. Putusan hakim dalam hal ini berupa kalimat kurang dimengerti oleh khalayak umum. Berdasarkan hasil tersebut peneliti ingin membantu khalayak umum/pembaca mengetahui bahasan yang berasal dari putusan hakim tersebut.

Salah satu perkembangan teknologi yang berkaitan dengan *text* dapat dimanfaatkan untuk mendapatkan topik dari kumpulan kalimat menjadikan fokus utama dalam penelitian ini. Berdasarkan permasalahan tersebut, penelitian ini akan melakukan pengelompokan (*clustering*) terhadap putusan hakim pada perkara di Pengadilan Negeri Sleman, D.I. Yogyakarta. Salah satu metode yang dapat digunakan untuk mengelompokkan hasil putusan hakim pada suatu perkara yaitu menggunakan analisis *topic modeling*. Algoritma dalam *topic modeling* ini dapat menemukan tema utama yang meliputi kumpulan dokumen yang besar dan tidak terstruktur. Oleh karena itu peneliti tertarik untuk menemukan tema dari putusan hakim di Pengadilan Negeri Sleman, D.I. Yogyakarta tahun 2016 – Januari 2020. Dalam melakukan analisis *topic modeling*, peneliti menggunakan metode *Latent Dirichlet Allocation* (LDA). LDA merupakan salah satu algoritma yang sering digunakan dalam analisis *topic modeling*. LDA digunakan untuk meringkas, melakukan *clustering*, menghubungkan maupun memproses data yang sangat besar karena LDA menghasilkan daftar topik yang diberi bobot untuk masing-masing

dokumen. Dengan menggunakan analisis tersebut, maka dapat diketahui topik-topik apa saja yang sering muncul, sehingga memudahkan dalam mengetahui dan memahami jenis perkara pidana yang sering terjadi di Kabupaten Sleman.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang di atas, dapat dirumuskan beberapa permasalahan sebagai berikut.

1. Bagaimana gambaran umum karakteristik pada data perkara pidana di Pengadilan Negeri Sleman pada tahun 2016 – Januari 2020?
2. Bagaimana hasil implementasi *topic modeling* dengan metode *Latent Dirichlet Allocation* (LDA) berdasarkan data putusan hakim pada perkara di Pengadilan Negeri Sleman tahun 2016 – Januari 2020?

1.3 Batasan Masalah

Untuk menghindari meluasnya pembahasan dalam penelitian ini, maka berikut merukan beberapa batasan masalah yang digunakan dalam penelitian ini.

1. Jenis acara pidana yang digunakan adalah pidana biasa yang telah berstatus minutasi.
2. Data yang digunakan diakses dan diambil pada *website* <http://pn-sleman.go.id/sipp/> paling lambat pada tanggal 14 Februari 2020.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, maka tujuan dari penelitian ini sebagai berikut.

1. Untuk mengetahui gambaran umum karakteristik pada data perkara pidana di Pengadilan Negeri Sleman pada tahun 2016 – Januari 2020.
2. Untuk mengetahui hasil implementasi *topic modeling* berdasarkan data putusan hakim pada perkara di Pengadilan Negeri Sleman pada tahun 2016 – Januari 2020.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut.

1. Menambah ilmu pengetahuan dan wawasan mengenai analisis *topic modeling*.
2. Penelitian ini diharapkan dapat memberikan informasi mengenai jenis perkara yang sering terjadi di Kabupaten Sleman pada tahun 2016 – Januari 2020 kepada aparat penegak hukum, admin website <http://pn-sleman.go.id/>, dan khalayak umum.
3. Dapat dijadikan sebagai referensi bagi pembaca untuk penelitian selanjutnya.



BAB 2 TINJAUAN PUSTAKA

Penelitian terdahulu digunakan sebagai bahan acuan dalam penulisan penelitian yang dilakukan berupa hasil penelitian sebelumnya maupun perkembangan dan penerapan terkini dari teori yang hendak diteliti. Terdapat beberapa penelitian terdahulu yang berkaitan dengan penelitian yang dilakukan oleh peneliti. Berikut merupakan penelitian yang menjadi acuan dalam penelitian ini.

Penelitian mengenai LDA yang pertama dilakukan oleh (Fitriasih & Kusumaningrum, 2019). Penelitian ini membahas tindak kejahatan *cyber-bullying* dan *cyber-harrassment* pada *tweet* di Twitter yang kemudian dikategorikan kedalam *crime indicated opinion* menggunakan metode LDA. Tujuan dari penelitian ini adalah mengklasifikasikan opini *tweet* yang kemudian dikembangkan kedalam *neural opinion* atau *crime indicated opinion*. Hasil dari penelitian menunjukkan bahwa model klasifikasi terbaik untuk 1000 iterasi dengan nilai akurasi sebesar 66%, nilai sensitivitas sebesar 77% dengan nilai hiperparameter *alpha* 0,00001 dan *beta* 0,1.

Penelitian mengenai LDA *topic modeling* yang kedua dilakukan oleh (ELvira, Cobo, Viedma, & Herrera, 2019). Penelitian ini membahas banyaknya sumber informasi bulletin resmi di Negara Spanyol yang dapat diakses namun hasil informasi yang diinginkan tidak direkomendasikan. Tujuan dari penelitian ini adalah membedakan dan memfilter informasi yang sesuai dengan preferensi dengan mengusulkan metode untuk melabeli dokumen secara otomatis berdasarkan LDA. Hasil dari penelitian ini menunjukkan bahwa dengan menggunakan pendekatan LDA sistem dapat merekomendasikan (pada sudut pandang teoretis) sebanyak 23% dibandingkan dengan sebelumnya yang hanya 11%.

Penelitian mengenai LDA *topic modeling* yang ketiga dilakukan oleh (Fajriyanto, 2018). Penelitian ini membahas penerapan metode Bayesian dalam model LDA di media sosial Twitter dengan *keyword* “@kompascom” dalam

rentang waktu 11-25 Mei 2018. Tujuan dari penelitian ini adalah mengetahui berita dominan yang dibahas oleh masyarakat pada periode tertentu. Hasil dari penelitian ini menunjukkan bahwa topik yang dominan dibahas masyarakat adalah penyebaran gambar atau video lokasi bom di Surabaya dapat ikut menyebarkan teror ketakutan yang menjadi tujuan pelaku bom yang diterbitkan oleh @kompascom pada tanggal 11 Mei 2018 sampai 25 Mei 2018.

Penelitian mengenai LDA *topic modeling* selanjutnya dilakukan oleh (Porter, 2018). Penelitian ini membahas pasar gelap *online* yang menjual barang-barang ilegal seperti obat-obatan, senjata api, dan *malware*. Tujuan dari penelitian ini adalah memahami dengan cepat pasar gelap *online* memengaruhi pasar dan pilihan alat yang digunakan oleh pengguna pasar *darknet* pada tanggal 5 November 2016 hingga 5 November 2017 dengan menentukan topik dan istilah yang menonjol, dan bagaimana mereka telah berubah selama satu tahun. Hasil dari penelitian menunjukkan bahwa *DarkNetMarkets* dapat mengungkapkan *ternd* dan kata kunci yang berkaitan dengan kegiatan kriminal dan sistem mereka. Pengguna *dark web* juga lebih berhati-hati tentang keamanan mereka karena penegakan hukum baru-baru ini.

Penelitian mengenai LDA *topic modeling* dilakukan oleh (Hariyady, Basuki, & Meidina, 2018). Penelitian ini membahas keluhan-keluhan pengguna jasa PT. Gojek Indonesia melalui media sosial twitter yang langsung ditujukan kepada akun resmi Gojek. Hasilnya diharapkan mampu membantu dalam pengembangan kinerja perusahaan. Tujuan dari penelitian ini adalah mendeteksi topik keluhan pengguna jasa Gojek menggunakan analisis *topic modeling*. Hasil dari penelitian menunjukkan bahwa topik keluhan pengguna jasa Gojek terdapat 5 topik. topik pertama cenderung membahas mengenai kesulitan dalam melakukan pemesanan, topik 2 yaitu aplikasi *error*, topik 3 yaitu kesulitan *log in* pada aplikasi, topik 4 yaitu respon perusahaan terhadap keluhan pelanggan, dan topik 5 yaitu pembatalan pemesanan oleh *driver*.

Penelitian mengenai tindak pidana dilakukan oleh (Maulana, 2017). Penelitian ini membahas tindak pidana di Kabupaten Sleman pada tahun 2013-2015. Tujuan dari penelitian ini adalah mengelompokkan kecamatan di Kabupaten

Sleman berdasarkan kasus tindak pidana di Kabupaten Sleman pada tahun 2013-2015 menggunakan metode *Self Organizing Maps* (SOM). Hasil dari penelitian menunjukkan bahwa tindak pidana yang paling menonjol di Kabupaten Sleman, yaitu pencurian, penipuan, penganiayaan, pencopetan, dan tindak asusila. Didapatkan hasil empat klaster dari 17 kecamatan di Kabupaten Sleman.

Bab ini dapat pula dirinci dalam bentuk tabel. Perbandingan penelitian terdahulu dengan penelitian yang penulis lakukan pada penulisan ini dapat dilihat pada tabel 2.1 berikut.

Tabel 2.1. Perbandingan dengan Penelitian Terdahulu

No	Penulis	Judul Penelitian	Metode	Persamaan	Perbedaan
1.	(Fitriasih & Kusumaningrum, 2019)	Analisis Klasifikasi Opini <i>Tweet</i> Pada Media Sosial Twitter Menggunakan <i>Latent Dirichlet Allocation</i> (LDA)	<i>Latent Dirichlet Allocation</i> (LDA) untuk Analisis Sentimen	Sama-sama menggunakan LDA	Studi kasus yang digunakan berbeda
2.	(ELvira, Cobo, Viedma, & Herrera, 2019)	<i>Latent Dirichlet Allocation (LDA) for Improving The Topic Modeling of The official Bulletin of The Spanish State (BOE)</i>	<i>Latent Dirichlet Allocation</i> (LDA)	Sama-sama menggunakan LDA	Studi kasus yang digunakan berbeda
3.	(Fajriyanto, 2018)	Penerapan Metode Bayesian dalam Model <i>Latent Dirichlet Allocation</i> di Media Sosial.	Bayesian dalam <i>Latent Dirichlet Allocation</i> (LDA)	Sama-sama menggunakan LDA	Studi kasus yang digunakan berbeda
4.	(Porter, 2018)	<i>Analyzing The DarkNetMarkets Subreddit for Evaluations of Tools</i>	LDA <i>Topic Modeling</i>	Sama-sama menggunakan LDA <i>Topic Modeling</i>	Studi kasus yang digunakan berbeda

No	Penulis	Judul Penelitian	Metode	Persamaan	Perbedaan
		<i>and Trends Using LDA Topic Modeling</i>			
5.	(Hariyady, Basuki, & Meidina, 2018)	Implementasi Algoritma Deteksi Topik Keluhan Pelanggan Jasa Ojek Online Berdasarkan Komentar Media Sosial	<i>Topic Modeling</i>	Sama-sama menggunakan <i>Topic Modeling</i>	Studi kasus yang digunakan berbeda
6.	(Maulana, 2017)	Aplikasi <i>Self Organizing Maps</i> (SOM) Untuk Pengelompokan Tindak Pidana Di Kabupaten Sleman Daerah Istimewa Yogyakarta (Studi Kasus : Jumlah Kasus Tindak Pidana Di Kabupaten Sleman, Daerah Istimewa Yogyakarta Tahun 2013-2015)	<i>Self Organizing Maps</i> (SOM)	Sama-sama menggunakan studi kasus mengenai tindak pidana di Kabupaten Sleman, D.I. Yogyakarta	Metode yang digunakan berbeda

Sehingga dalam penelitian ini, penulis menggunakan metode *Latent Dirichlet Allocation* (LDA) dengan studi kasus data putusan hakim pada perkara di Pengadilan Negeri Sleman tahun 2016 – Januari 2020.

BAB 3 LANDASAN TEORI

3.1 Kejahatan

Kejahatan dalam definisi yuridis berbeda dengan definisi kejahatan dalam kriminologi yang dipandang secara sosiologis. Secara yuridis, kejahatan merupakan suatu tindakan yang melanggar undang-undang atau ketentuan yang berlaku dan diakui secara legal. Secara kriminologi yang dipandang sosiologis, kejahatan merupakan suatu pola tingkah laku yang merugikan masyarakat (terdapat korban) dan suatu pola tingkah laku yang mendapatkan reaksi sosial dari masyarakat (Mursalin, 2012).

Kejahatan merupakan tindakan yang mengandung suatu *onrecht* sehingga orang memandang perilaku tersebut pantas dihukum meskipun tidak dicantumkan dalam undang-undang sebagai perbuatan terlarang oleh pembuat undang-undang. Berbeda dengan kejahatan, pelanggaran merupakan orang yang umumnya baru mengetahui tindakan tersebut merupakan pelanggaran yang bersifat melawan hukum sehingga dapat dihukum yaitu setelah tindakan tersebut dinyatakan dilarang dalam undang-undang (Hasanah, 2017).

3.2 Tindak Pidana

Tindak pidana merupakan terjemahan dari *strafbaarfeit* tapi dalam Kitab Undang-undang Hukum Pidana (KUHP) tidak memberikan penjelasan maksud dari *strafbaarfeit* itu sendiri. Dalam bahasa Belanda *strafbaarfeit* dibagi menjadi dua unsur pembentuk kata, yaitu *strafbaar* dan *feit*. Kata *feit* diartikan sebagian dari kenyataan, sedangkan kata *strafbaar* berarti dapat dihukum. Dari dua unsur pembentuk kata tersebut sehingga secara harafiah perkataan *strafbaarfeit* diartikan sebagai sebagian dari kenyataan yang dapat dihukum (Hartanti, 2009).

Menurut Simons dalam (Hartanti, 2009) *strafbaarfeit* adalah tindakan melanggar hukum yang dilakukan dengan sengaja maupun tidak sengaja oleh seseorang yang dapat dipertanggungjawabkan atas tindakannya dan oleh undang-undang telah dinyatakan sebagai tindakan yang dapat dihukum. Sedangkan menurut

Prasetyo (2015) tindak pidana dianonimkan dengan kata delik (*delictum*) yang berasal dari bahasa Latin. Menurut Kamus Besar Bahasa Indonesia dalam (Prasetyo, 2015) delik merupakan perbuatan yang melanggar undang-undang tindak pidana sehingga dikenakan hukuman.

3.2.1 Pertanggungjawaban Pidana

Pertanggungjawaban pidana dalam istilah asing disebut teorekenbaardheid atau responsibility yang menjurus kepada pembedaan pelaku dengan maksud untuk menentukan apakah seorang terdakwa atau tersangka dipertanggungjawabkan atas suatu tindakan pidana yang terjadi atau tidak. Seseorang akan dipertanggungjawabkan atas tindakannya apabila tindakan tersebut melawan hukum serta tidak ada alasan pembenar atau peniadaan sifat melawan hukum untuk pidana yang dilakukannya, hal tersebut dilihat dari sudut terjadinya tindakan yang dilarang. Sedangkan jika dilihat dari sudut kemampuan bertanggungjawab maka hanya seseorang yang mampu bertanggungjawab yang dapat dipertanggungjawabkan atas perbuatannya (Ilyas, 2012).

3.2.2 Jenis-jenis Pidana

Dalam Kitab Undang-Undang Hukum Pidana (KUHP) hukuman/pidana terdapat dalam Pasal 10 KUHP yang mengatakan bahwa pidana terdiri atas:

1. Pidana pokok, yaitu:
 - a. Pidana mati, pelaksanaan pidana mati harus dengan Keputusan Presiden sekalipun terpidana menolak untuk memohon pengampunan atau grasi dari Presiden. Sedangkan penundaan pidana mati jika terpidana sakit jiwa atau wanita yang sedang hamil (Ilyas, 2012).
 - b. Pidana penjara, menurut P.A.F. Lamintang dalam (Ilyas, 2012) yaitu pembatasan kebebasan bergerak dari terpidana, yaitu dilakukan dengan menutup orang tersebut dalam sebuah Lembaga Perasyarakatan dengan mewajibkan orang itu untuk mentaati semua peraturan tata tertib bagi mereka yang telah melanggar peraturan tersebut.
 - c. Pidana kurungan, jangka waktu pidana kurungan lebih ringan dibandingkan dengan pidana penjara seperti yang tertuang dalam pasal

69 ayat (1) KUHP, bahwa berat ringannya pidana ditentukan oleh urutan dalam Pasal 10 KUHP. Lama hukuman pidana kurungan yaitu sekurang-kurangnya satu hari dan paling lama satu tahun seperti yang tertuang dalam pasal 18 KUHP (Ilyas, 2012).

- d. Pidana denda, merupakan kewajiban seorang yang dijatuhi pidana denda tersebut oleh Hakim/Pengadilan karena orang tersebut telah melakukan suatu perbuatan yang dapat dipidana. Pidana denda dapat dibayar oleh orang lain selama atas nama terpidana (Ilyas, 2012).
 - e. Pidana tutupan, seperti yang tertuang dalam UU No. 20 Tahun 1946 Pasal 1 pidana tambahan yaitu pidana yang menggantikan hukuman penjara dalam hal tersebut dalam Pasal 2.
2. Pidana tambahan, yaitu:
- a. Pencabutan hak-hak tertentu, ketentuan hak-hak yang dapat dicabut oleh hakim dengan suatu putusan pengadilan tertuang dalam Pasal 35 ayat (1) KUHP. Sedangkan dalam hal dilakukannya lamanya pencabutan hak diatur oleh hakim yang tertuang dalam Pasal 38 ayat (1) KUHP. Dalam hal ini hakim tidak berwenang seorang dari jabatannya jika dalam aturan-aturan khusus ditentukan penguasa lain untuk pemecatan itu (Ilyas, 2012).
 - b. Perampasan barang-barang tertentu, merupakan jenis pidana harta kekayaan dengan ketentuannya yang tertuang dalam pasal 39 KUHP. Perampasan barang-barang yang tidak disita sebelumnya diganti pidana kurungan yang paling sedikit satu hari dan paling lama enam bulan apabila barang-barang tersebut tidak diserahkan atau harganya menurut taksiran dalam putusan hakim tidak dibayar. Namun kurungan pengganti dapat dihapuskan apabila barang-barang yang dirampas diserahkan (Ilyas, 2012).
 - c. Pengumuman putusan hakim, dimaksudkan untuk pencegahan supaya masyarakat terhindar dari kelihaiian busuk atau kecerobohan seorang pelaku. Pidana tambahan ini hanya berlaku apabila secara tegas

ditentukan berlaku untuk pasal-pasal tindak pidana tertentu (Ilyas, 2012).

3.2.3 Acara Pemeriksaan

Berikut merupakan jenis-jenis acara pemeriksaan di sidang pengadilan menurut UU No. 8 Tahun 1981 KUHAP:

1. Acara Pemeriksaan Biasa

Berdasarkan Pasal 152-202 KUHAP, acara pemeriksaan biasa diawali dengan pemeriksaan persiapan dengan majelis hakim 3 orang dengan tahapan penanganan seketanya yang pertama prosedur dismissal yaitu pemeriksaan administrative untuk menetapkan suatu gugatan dapat diterima atau tidak, kedua pemeriksaan persiapan yaitu untuk melengkapi gugatan yang kurang jelas, dan yang ketiga pemeriksaan di sidang pengadilan. Bentuk akhir dari acara pemeriksaan biasa adalah putusan (vonis).

2. Acara pemeriksaan Singkat

Berdasarkan Pasal 203-204 KUHAP, acara pemeriksaan singkat dilakukan terhadap perlawanan yang diajukan terhadap penetapan dari prosedur dismissal dalam tenggang waktu 14 hari setelah penetapan diucapkan. Pemeriksaan singkat ini dilakukan karena adanya perlawanan penggugat tentang gugatannya yang tidak diterima atau tidak mendasar. Penundaan pelaksanaan TUN tidak untuk menyelesaikan pokok sengketa dengan bentuk akhirnya berupa penetapan.

3. Acara pemeriksaan Cepat

Berdasarkan Pasal 205-216 KUHAP, acara pemeriksaan cepat dilakukan karena kepentingan mendesak yang harus dapat disimpulkan dengan hakim tunggal. Permohonan dikabulkan atau tidaknya dalam jangka waktu 14 hari setelah diterimanya permohonan pemeriksaan acara cepat oleh ketua pengadilan. Kemudian ketua pengadilan dalam jangka waktu 7 hari setelah dikeluarkannya penetapan menentukan hari, tempat, dan waktu sidang dilakukan tanpa melalui prosedur pemeriksaan persiapan. Sedangkan tenggang waktu jawaban dan pembuktian bagi kedua belah pihak masing-

masing tidak melebihi 14 hari dengan bentuk akhir dari acara pemeriksaan cepat adalah putusan (vonis).

3.3 Hukum Pidana

Hukum pidana dalam artian obyektif menurut Mezger adalah aturan hukum yang mengikat kepada suatu perbuatan yang memenuhi syarat tertentu suatu akibat yang berupa pidana. Sedangkan hukum pidana dalam artian subyektif dapat diartikan secara luas dan sempit. Dalam artian luas diartikan sebagai hak dari negara untuk mengenakan atau mengancam pidana terhadap perbuatan tertentu. Dalam artian sempit diartikan sebagai hak untuk menuntut perkara-perkara pidana, menjatuhkan, dan melaksanakan pidana terhadap orang yang melakukan perbuatan dilarang (oleh hukum pidana) atau juga disebut tindak pidana (Usfa & Tongat, 2004).

3.4 Statistika Deskriptif

Menurut Iqbal Hasan dalam (Astuti, 2014) statistika deskriptif merupakan bagian dari ilmu statistika yang mempelajari cara mengumpulkan dan menyajikan data sehingga mudah dipahami. Statistika deskriptif berfungsi menerangkan keadaan, gejala, atau persoalan dengan kata lain hanya gambaran secara umum dari data yang didapatkan. Dalam penulisan ini, gambaran umum yang digunakan didapatkan melalui *pie chart* dan *bar chart*.

3.5 Artificial Intelligence

Menurut (Wijaya, 2013) *Artificial Intelligence* (AI) atau yang sering disebut dengan kecerdasan buatan merupakan bagian ilmu komputer yang membuat suatu komputer (mesin) menjadi cerdas sehingga dapat menggantikan atau menirukan pekerjaan yang dilakukan oleh manusia. Awalnya komputer diciptakan hanya sebagai alat hitung, namun seiring dengan berkembangnya zaman komputer diharapkan dapat digunakan untuk mengerjakan pekerjaan yang dilakukan oleh manusia. Dalam menjadikan komputer yang cerdas maka komputer perlu ditambahkan komponen pengetahuan dan kemampuan menalar. Berikut merupakan

dua bagian utama yang dibutuhkan dalam penerapan konsep kecerdasan buatan pada komputer (Andayu, 2013).

1. Basis pengetahuan (*knowledge base*), yang berisi fakta, teori, dan pemikiran yang berhubungan satu sama lain.
2. Motor inferensi (*inference engine*), kemampuan menarik kesimpulan berdasarkan pengetahuan.



Gambar 3.1. Penerapan Konsep Kecerdasan Buatan di Komputer

Sumber: (Andayu, 2013)

Menurut (Andayu, 2013) lingkup utama dalam kecerdasan buatan sebagai berikut.

1. Sistem Pakar (*Expert System*), komputer digunakan sebagai sarana penyimpanan pengetahuan para pakar sehingga komputer mempunyai keahlian dalam menyelesaikan permasalahan dengan meniru keahlian yang dimiliki oleh pakar.
2. Pemrosesan Bahasa Alami (*Natural Language Processing*), pengguna dapat berkomunikasi dengan komputer menggunakan bahasa yang mereka gunakan sehari-hari.
3. Pengenalan Ucapan (*Speech Recognition*) sehingga manusia dapat berkomunikasi dengan komputer menggunakan suara.
4. Robotika Sistem Sensor (*Robotics and Sensory Systems*), contoh bagian robotika pada sistem sensor seperti sistem visi dan pencitraan serta sistem pengolahan sinyal.
5. *Computer Vision*, untuk menerjemahkan suatu citra ke dalam komputer.
6. *Intelligent Computer-aided Instruction*, komputer dimanfaatkan sebagai tutor yang dapat melatih dan mengajar.
7. Mesin belajar (*Machine Learning*), merupakan kumpulan metode dalam pemecahan masalah yang digunakan untuk melatih komputer dengan menganalisa kasus-kasus yang telah terjadi.

3.6 *Machine Learning*

Machine learning (ML) merupakan pendekatan AI yang banyak digunakan dalam menggantikan atau menirukan perilaku manusia untuk menyelesaikan masalah atau melakukan otomatisasi. Agar dapat menggantikan atau menirukan perilaku manusia, sistem dalam komputer atau mesin harus memiliki pemampuan untuk belajar. ML mengharuskan adanya proses pelatihan, pembelajaran, ataupun *training*, sehingga ML membutuhkan data untuk dipelajari yang disebut *training* (Ahmad, 2017). *Machine learning* terbagi menjadi tiga bagian yaitu (Rauhan, 2019):

1. *Supervised learning*, pada algoritma ini sistem diberikan *training data set* berupa informasi *input* dan *output* yang diinginkan dikarenakan algoritma ini bersifat prediktif sehingga sistem akan mempelajari berdasarkan data yang ada. Sistem akan mencari pola dari data set sehingga pola tersebut akan dijadikannya sebagai acuan untuk data-data berikutnya.
2. *Unsupervised learning*, algoritma ini tidak diberikan *training data set* sehingga membutuhkan pembelajaran dari data yang ada. algoritma ini bersifat deskriptif, sehingga berguna untuk mengelompokkan atau mengkategorikan data.
3. *Reinforcement learning*, ide dasar pada algoritma ini yaitu terdapat *agent* yang ditempatkan pada sebuah *environment* yang tempatnya tidak diketahui, kemudian agent akan mengeksplorasi seluruh *environment* sehingga didapatkan istilah *reward* dan *error*. Apabila agent menemukan sebuah *error* maka ia akan terus mencari jalan hingga mendapatkan *reward*. Eksplorasi ini beroperasi seperti *trial and error*.

3.7 *Text Mining*

Menurut Hikmah (2018) *text mining* merupakan istilah untuk penggalian suatu informasi berupa teks yang sumber datanya didapatkan dari dokumen. Menurut Feldman dalam (Hikmah, 2018) tujuan dari *text mining* yaitu untuk menemukan kata-kata yang dapat mewakili isi dokumen sehingga dapat dicari hubungan antar dokumen yang lain. *Text mining* biasanya meliputi proses penataan

teks *input*, menemukan pola dalam data terstruktur, dan evaluasi serta interpretasi *output*. Tugas umum dari *text mining* meliputi kategorisasi teks, pengelompokan teks, ringkasan dokumen, ekstraksi kata kunci, dan lain-lain (Tong & Zhang, 2016).

3.12 *Text Preprocessing*

Text preprocessing merupakan langkah penting dalam *Natural Language Processing* (NLP). *Text preprocessing* mengubah teks menjadi lebih sederhana sehingga algoritma *machine learning* dapat bekerja lebih baik. *Text preprocessing* merupakan tindakan menghilangkan karakter-karakter tertentu yang terkandung dalam dokumen (Siregar, Sinaga, & Arianto, 2017). Berikut merupakan tahapan proses *text preprocessing* yang digunakan pada penelitian ini.

a. *Lower Casing*

Lower casing adalah proses mengubah huruf yang ada di kumpulan teks atau dokumen menjadi huruf kecil semua.

b. *Remove Punctuation*

Remove punctuation merupakan proses penghapusan tanda baca pada suatu dokumen atau kumpulan teks. Pada *text preprocessing* tanda baca seperti [! " # \$ % & ' () * + , - . / : ; < = > ? @ [\] ^ _ ` { | } ~] dalam kalimat tidak memiliki pengaruh.

c. *Stopword Removal*

Stopword merupakan kata umum yang biasa muncul dan dianggap tidak memiliki makna. Jadi *stopword removal* adalah tahapan mengambil kata-kata yang dianggap penting saja menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting).

d. *Tokenizing*

Tokenizing merupakan proses pemisahan dokumen atau kumpulan teks menjadi potongan-potongan token (kata, angka, simbol, tanda baca, dan entitas penting lainnya).

3.13 *Pembobotan Kata (Term Weighting)*

Term frequency-Inverse Document Frequency (TF-IDF) merupakan seberapa pentingnya setiap kata dalam dokumen yang dilakukan dengan pembobotan kata

tersebut. TF-IDF sering digunakan dalam pembobotan text mining. Menurut Nurjannah dkk (2013) TF-IDF dilakukan dengan mengalikan dua metrik untuk perhitungan bobot, yaitu berapa kali kemunculan sebuah kata di dalam dokumen dan inverse frekuensi dokumen yang mengandung kata tersebut. Frekuensi kemunculan kata yang semakin sering muncul di dalam dokumen menunjukkan bahwa bobotnya semakin besar dan seberapa penting kata tersebut. Jika kata tersebut sering muncul dalam banyak dokumen menunjukkan bahwa bobotnya semakin kecil dan seberapa umum kata tersebut.

TF digunakan untuk menghitung bobot kata yang muncul dalam sebuah dokumen. Menurut Siregar dkk (2017) berikut merupakan empat cara yang dapat digunakan untuk menghitung nilai TF.

- a. *Raw TF*, dalam memperoleh nilai TF dihitung *term* berdasarkan jumlah kemunculan *term* tersebut dalam dokumen.
- b. *Logarithmic TF*, nilai TF diberikan dengan menggunakan fungsi logaritmik dalam matematika. Dalam perhitungan ini untuk menghindari dominansi dokumen yang mengandung sedikit *term* dalam *query* namun mempunyai frekuensi yang tinggi. Berikut merupakan rumus *logarithmic TF*.

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

Dimana nilai TF merupakan frekuensi *term* (*t*) pada *document* (*d*).

- c. *Binary TF*, TF bernilai 0 apabila tidak ada *term* pada dokumen dan bernilai 1 apabila ada *term* dalam dokumen. Banyaknya kemunculan *term* pada dokumen tidak berpengaruh.
- d. *Augmented TF*, nilai TF diberikan frekuensi muncul *term* pada sebuah dokumen. Berikut merupakan rumus *augmented TF*.

$$TF = 0.5 + 0.5x \left[\frac{TF}{\max TF} \right] \quad (3.2)$$

Dimana nilai *max TF* merupakan jumlah muncul *term* terbanyak pada dokumen yang sama.

IDF digunakan untuk menunjukkan hubungan ketersediaan *term* dalam sebuah dokumen. Nilai IDF akan semakin besar apabila jumlah dokumen yang mengandung *term* (*t*) semakin sedikit. Berikut merupakan rumus dari IDF.

$$IDF_t = \ln\left(\frac{D}{DF_t}\right) + 1 \quad (3.3)$$

Dimana:

IDF = *Inverse Document Frequency*

D = Total dokumen

DF_t = frekuensi dokumen dari *term t*

Rumus yang digunakan dalam perhitungan TF pada penelitian ini yaitu *raw TF*.

Sehingga rumus bobot *term* dihitung menggunakan rumus sebagai berikut.

$$W_{dt} = TF_{dt} \times IDF_t \quad (3.4)$$

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

Dimana:

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

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

D = Total dokumen

IDF_{dt} = *Inverse document frequency*

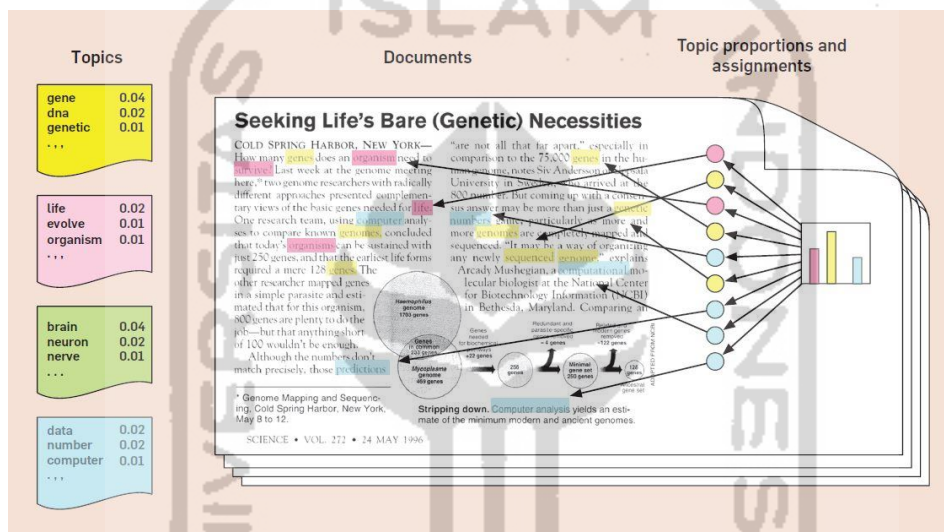
Berapapun nilai TF_{dt}, apabila D = DF_t, maka hasilnya nol karena merupakan hasil dari ln 1 untuk perhitungan IDF. Ditambahkan nilai satu pada sisi IDF, sehingga perhitungan pembobotannya sebagai berikut.

$$W_{dt} = TF_{dt} \times \left(\ln\left(\frac{D}{DF_t}\right) + 1 \right) \quad (3.5)$$

3.14 Topic Modeling

Menurut (Jing, 2014) *Topic modeling* adalah algoritma untuk menemukan tema utama yang meliputi kumpulan dokumen yang besar dan tidak terstruktur. Dalam *machine learning*, *topic modeling* dapat dianggap sebagai metode *unsupervised learning* yang digunakan dalam model generatif yang menyediakan kerangka kerja probababilistik. *Topic modeling* menyediakan serangkaian algoritma untuk menemukan struktur tematis tersembunyi dalam koleksi teks yang besar. Hasil dari algoritma *topic modeling* dapat digunakan untuk meringkas, memvisualisasikan, mengeksplorasi, dan berteori tentang *corpus* (Blei, 2013).

Pada dasarnya *topic modeling* menurut Blei adalah bahwa sebuah topik terdiri dari kata-kata tertentu yang menyusun topik tersebut, dan dalam satu dokumen kemungkinan memiliki beberapa topik dengan probabilitas masing-masing. Dokumen-dokumen merupakan objek yang dapat diamati, sedangkan topik, distribusi topik setiap dokumen merupakan struktur tersembunyi, sehingga topic modeling bertujuan untuk menemukan topik dan kata-kata yang terdapat dalam topik tersebut (Putra, 2017). Konsep *topic modeling* menurut Blei, dapat dilihat pada gambar 3.2 berikut.



Gambar 3.2. Representasi Model LDA

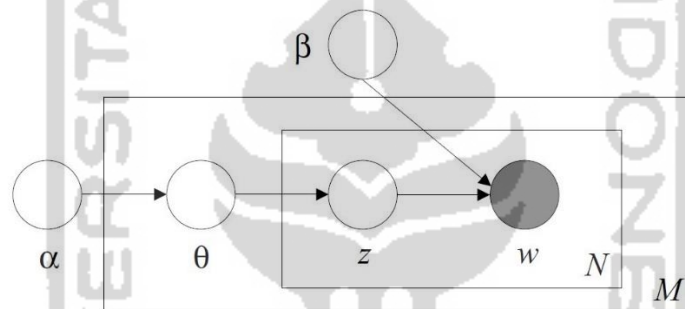
Sumber: (Blei, 2012)

Salah satu cara untuk berpikir tentang bagaimana proses *topic modeling* bekerja adalah dengan membayangkan bekerja melalui artikel dengan kumpulan highlighters. Dalam gambar 3.2, Blei menyalin kata-kata yang dikelompokkan berdasarkan warna yang telah ditetapkan sendiri. Daftar kata tersebut adalah topik, dan setiap warna mewakili topik yang berbeda (Brett, 2012).

3.15 Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) merupakan salah satu algoritma yang sering digunakan dalam *topic modeling*. Menurut (Blei, Ng, & Jordan, Latent Dirichlet Allocation, 2003), LDA merupakan model *probabilistic generative* dari sebuah *corpus*. Dalam LDA, kumpulan dokumen memperlihatkan banyak topik, dimana setiap dokumen dianggap sebagai campuran dari berbagai topik. Kumpulan

dokumen memiliki distribusi probabilitas topik, artinya mayoritas dokumen membicarakan tentang topik yang sama (Jing, 2014). Campbell, Hindle, dan Stroulia (2014) dalam penulisan (Putra, 2017) LDA dapat digunakan untuk meringkas, melakukan *clustering*, menghubungkan maupun memproses data yang sangat besar karena LDA menghasilkan daftar topik yang diberi bobot untuk masing-masing dokumen. Distribusi yang digunakan untuk mendapatkan distribusi topik setiap dokumen disebut distribusi *Dirichlet*, kemudian hasil dari *Dirichlet* digunakan untuk mengalokasikan kata-kata pada dokumen untuk topik yang berbeda. Blei merepresentasikan model LDA secara visual seperti pada gambar 3.3 berikut.



Gambar 3.3. Representasi Model LDA

Sumber : (Blei, Ng, & Jordan, Latent Dirichlet Allocation, 2003)

Model LDA direpresentasikan sebagai model grafis probabilistik seperti pada gambar 3.3. Sesuai visualisasi pada gambar 3.3 di atas, terdapat tiga tingkatan pada LDA *modeling*. Parameter α dan β merupakan parameter distribusi topik yang berada pada tingkatan *corpus*, yaitu kumpulan dari M dokumen. Parameter α digunakan dalam menentukan distribusi topik dalam dokumen, semakin besar nilai α dalam suatu dokumen, menandakan campuran topik yang dibahas dalam dokumen semakin banyak. Parameter β digunakan untuk menentukan distribusi kata dalam topik. Semakin tinggi nilai β , maka semakin banyak kata-kata yang ada di dalam topik, sedangkan semakin kecil nilai β , maka semakin sedikit kata-kata yang ada dalam topik sehingga topik tersebut mengandung kata-kata yang lebih spesifik. Variabel θ_m adalah variabel yang berada di tingkat dokumen (M). Variabel θ_m merepresentasikan distribusi topik untuk dokumen tertentu. Semakin tinggi nilai

θ , maka semakin banyak topik yang ada di dalam dokumen, sedangkan semakin kecil nilai θ , maka dapat dikatakan dokumen tersebut semakin spesifik pada topik tertentu. Variabel Z_n dan W_n adalah variabel tingkat kata (N). Variabel Z merepresentasikan topik dari kata tertentu pada sebuah dokumen sedangkan variabel W merepresentasikan kata yang berkaitan dengan topik tertentu yang terdapat dalam dokumen. Bentuk lingkaran merepresentasikan individual kata. Lingkaran berwarna abu-abu merepresentasikan variabel yang di *observe*, sedangkan lingkaran yang kosong merepresentasikan variabel laten atau variabel yang tidak secara langsung di *observe* (Blei, Ng, & Jordan, Latent Dirichlet Allocation, 2003). Berdasarkan notasi-notasi tersebut, berikut merupakan persamaan perhitungan probabilitas sebuah *corpus*.

$$p(D | \alpha, \beta) = \prod_{d=1}^M \int p(\theta_d | \alpha) \left(\prod_{n=1}^{N_d} \sum_{z_n} p(z_{dn} | \theta_d) p(w_{dn} | z_{dn}, \beta) \right) d\theta_d$$

3.16 Topic Coherence

Topic coherence adalah salah satu ukuran yang digunakan untuk mengevaluasi *topic modeling*. Ukuran *topic coherence* menilai satu topik dengan mengukur tingkat kesamaan semantik antara kata-kata dengan skor tingi dalam topik (Stevens, Kegelmeyer, Andrzejewski, & Buttler, 2012). *Coherence score* adalah skor yang menghitung jika kata-kata dalam topik yang sama masuk akal ketika disatukan. Semakin tinggi skor untuk jumlah k topik tertentu, artinya untuk setiap topik, akan ada lebih banyak kata yang terkait bersama dan topik tersebut akan lebih masuk akal (Tang, 2019).

BAB 4 METODOLOGI PENELITIAN

4.1 Populasi dan Sampling Penelitian

Populasi yang digunakan dalam penelitian ini adalah data perkara pidana yang terdapat di Pengadilan Negeri Sleman. Sedangkan data yang digunakan dalam penelitian ini adalah perkara pidana yang ada di Pengadilan Negeri Sleman pada tahun 2016 – Januari 2020.

4.2 Jenis dan Sumber Data

Jenis data yang digunakan dalam penelitian ini adalah data sekunder, yaitu data sistem informasi penelusuran perkara Pengadilan Negeri Sleman. Data tersebut diperoleh melalui *website* <http://pn-sleman.go.id/sipp/>.

4.3 Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah perkara pidana di Pengadilan Negeri Sleman pada tahun 2016 – Januari 2020. Devinisi variabel tersebut dapat dilihat pada tabel 4.1 berikut.

Tabel 4.1. Definisi Operasional Variabel

No.	Variabel	Definisi Operasioal Variabel
1	Putusan	Putusan pengadilan adalah pernyataan hakim yang diucapkan dalam sidang pengadilan terbuka, yang dapat berupa pemidanaan atau bebas atau lepas dari suatu tuntutan hukum dalam UU pasal 1 butir 11 KUHAP. Semua putusan pengadilan hanya sah dan memiliki kekuatan hukum jika diucapkan di sidang terbuka untuk umum (Effendi, 2016).
2	Klasifikasi Perkara	Jenis-jenis perkara setelah putusan hakim yang ada pada data perkara tindak pidana.
3	Minutasi	Proses yang dilakukan oleh panitera pengadilan dalam menyelesaikan proses administrasi yang terdiri dari

		pengetikan, pembendelan serta pengesahan pada suatu perkara
--	--	---

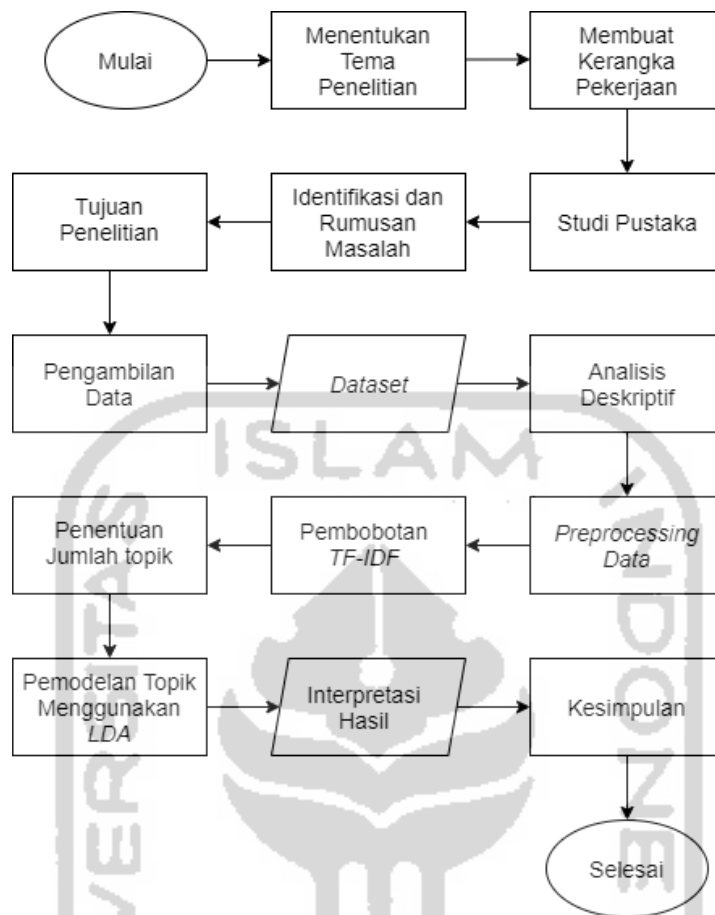
4.4 Metode Analisis Data

Software yang digunakan dalam penelitian ini adalah *Microsoft Excel 2013* dan *Anaconda3 5.3.1*. Metode yang digunakan dalam penelitian ini adalah sebagai berikut.

1. Analisis deskriptif, digunakan untuk mengetahui gambaran umum dari data tindak pidana yang ada di Pengadilan Negeri Sleman pada tahun 2016 – Januari 2020.
2. *Preprocessing data*, digunakan untuk mengubah teks menjadi lebih sederhana yang dapat mewakili dokumen.
3. Pembobotan kata menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) digunakan untuk pembobotan kata yang awalnya data teks menjadi data numerik.
4. Penentuan jumlah topik menggunakan *topic coherence*.
5. *Topic modeling* menggunakan algoritma *Latent Dirichlet Allocation* (LDA), digunakan untuk mengelompokkan setiap kata menjadi topik-topik dari setiap dokumen serta mencari hubungan antar katanya.

4.5 Tahapan Penelitian

Langkah atau tahapan pada penelitian ini dapat dilihat pada *flowchart* gambar 4.1 di bawah ini.



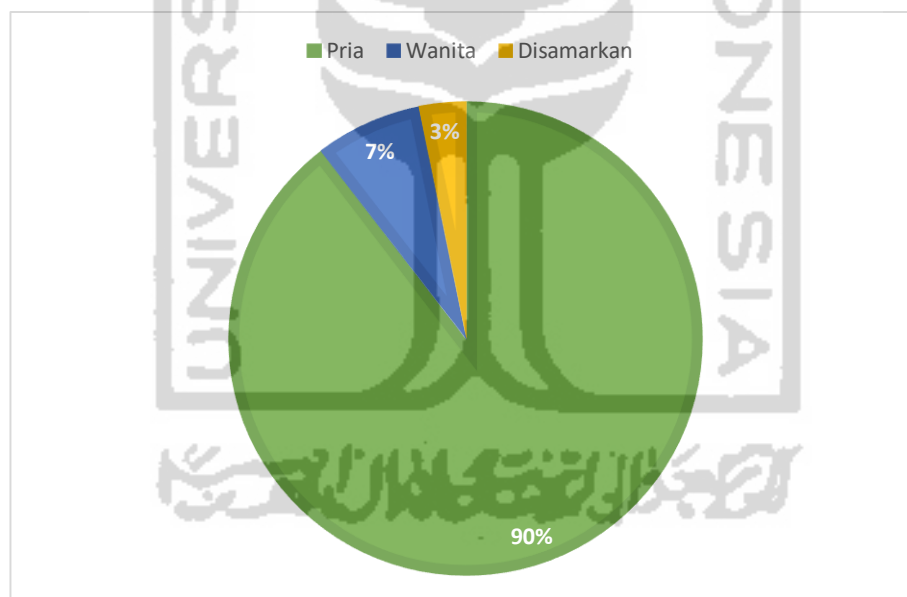
Gambar 4.1. *Flowchart Tahapan Penelitian*

BAB 5 HASIL DAN PEMBAHASAN

Pada bab ini akan dibahas mengenai *topic modeling* dengan menggunakan metode *Latent Dirichlet Allocation* (LDA) dengan data yang digunakan adalah data perkara pidana yang ada di Pengadilan Negeri Sleman pada tahun 2016 – Januari 2020. Sebelum ke tahap *topic modeling*, penulis melakukan *preprocessing data* dan pembobotan kata dengan TF-IDF.

5.1 Analisis Deskriptif

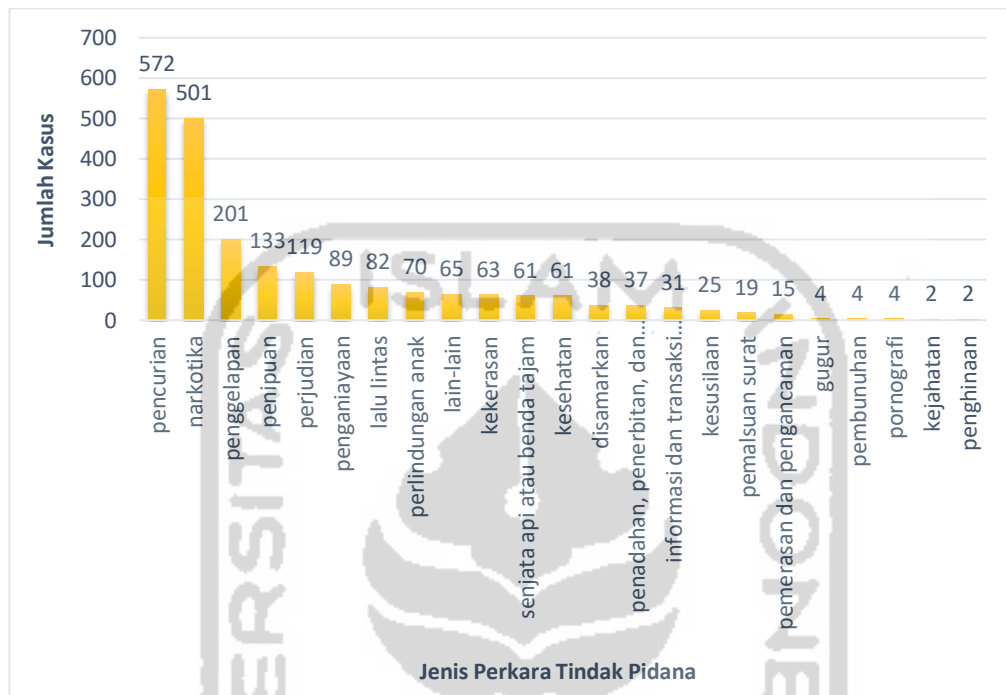
Analisis deskriptif dalam penulisan ini bertujuan untuk melihat gambaran umum data yang digunakan yaitu data pada perkara pidana yang ada di Pengadilan Negeri Sleman tahun 2016 – Januari 2020 yang diambil melalui *website* <https://pn-sleman.go.id/sipp>.



Gambar 5.1. Persentase Jenis Kelamin Tindak Pidana

Gambar 5.1 merupakan grafik mengenai jenis kelamin yang melakukan tindak kejahatan, dari 2198 perkara terdapat 2731 jiwa yang ikut serta dalam melakukan tindak kejahatan mereka. Sebanyak 2447 orang yang melakukan kejahatan di wilayah Kabupaten Sleman pada tahun 2016 – Januari 2020 berjenis kelamin pria atau sekitar 90%, sebanyak 198 orang yang melakukan tindak kejahatan adalah berjenis kelamin wanita atau sekitar 7%, sedangkan sisanya

sebanyak 86 orang atau sebanyak 3% yang melakukan tindak kejahatan namun tidak diketahui jenis kelaminnya yang artinya berarti bahwa perkara tersebut biasanya disamarkan/tidak dipublikasikan.



Gambar 5.2. Jumlah Perkara Tindak Pidana Berdasarkan Klasifikasi Perkara Setelah Putusan

Gambar 5.2 merupakan grafik mengenai jumlah perkara tindak pidana berdasarkan klasifikasi perkaranya setelah putusan hakim. Pada grafik tersebut menunjukkan secara runtut jumlah perkara pada klasifikasi perkara di wilayah Kabupaten Sleman pada tahun 2016 – Januari 2020. Dari gambar 5.2 tersebut dapat dilihat bahwa jumlah perkara tindak pidana berdasarkan klasifikasi perkara lima teratas yang paling menonjol di Kabupaten Sleman pada tahun 2016 – Januari 2020 adalah yang pertama pada grafik pencurian dengan jumlah kasus sebanyak 572 kasus, kedua pada grafik narkotika dengan jumlah kasus sebanyak 501 kasus, yang ketiga pada grafik penggelapan dengan jumlah kasus sebanyak 201 kasus, yang keempat pada grafik penipuan dengan jumlah kasus sebanyak 133 kasus, dan yang kelima pada grafik perjudian dengan jumlah kasus sebanyak 119 kasus.

5.2 Preprocessing Data

Langkah awal sebelum memulai tahap pengelompokan yaitu *preprocessing data*. *Preprocessing data* bertujuan untuk menghilangkan atau membersihkan teks pada data yang tidak diperlukan. Berikut merupakan contoh data yang digunakan dalam penelitian ini, sedangkan untuk data lengkap dapat dilihat dalam Lampiran.

Tabel 5.1 Data Awal Penelitian

Putusan
MENGADILI: Menyatakan Terdakwa Pulung Aji Wibowo Bin Hudi Sumarno Alm telah terbukti secara sah dan meyakinkan bersalah melakukan tindak pidana bersama-sama melakukan penyalahgunaan narkoba golongan I bagi diri sendiri; menjatuhkan pidana kepada Terdakwa oleh karena itu dengan pidana penjara selama 8 (delapan) bulan; Menetapkan masa penangkapan dan penahanan yang telah dijalani Terdakwa dikurangkan seluruhnya dari pidana yang dijatuhkan; Menetapkan agar Terdakwa tetap berada dalam tahanan; Menetapkan barang bukti berupa: 1 (satu) buah HP I Phone 6 warna gold dengan No. Simcard 0822 2172 2002 yang masih diperlukan sebagai barang bukti dalam perkara No 509/Pid.Sus/2019/PN Smn, maka dikembalikan kepada Penuntut Umum untuk dijadikan barang bukti dalam perkara No 509/Pid.Sus/2019/PN Smn; Membebaskan kepada Terdakwa membayar biaya perkara sejumlah Rp 2000 (dua ribu rupiah);

Tabel 5.1 di atas dapat digunakan untuk contoh *preprocessing data* dalam penulisan ini. Tahap yang dilakukan oleh peneliti dalam melakukan *preprocessing data* ini yaitu *lower casing*, *remove punctuation*, *stopword removal*, dan *tokenizing*.

a. Lower Casing

Lower casing digunakan untuk menyamakan teks dengan cara mengubah huruf capital menjadi huruf kecil. Hasil dari proses *lower casing* dapat dilihat pada tabel 5.2 berikut.

Tabel 5.2 Hasil *Lower Casing*

Sebelum	Sesudah
<p><u>MENGADILI</u>: Menyatakan <u>Terdakwa Pulung Aji Wibowo Bin Hudi Sumarno Alm</u> telah terbukti secara sah dan meyakinkan bersalah melakukan tindak pidana bersama-sama melakukan penyalahgunaan narkotika golongan <u>I</u> bagi diri sendiri; menjatuhkan pidana kepada <u>Terdakwa</u> oleh karena itu dengan pidana penjara selama 8 (delapan) bulan; <u>Menetapkan</u> masa penangkapan dan penahanan yang telah dijalani <u>Terdakwa</u> dikurangkan seluruhnya dari pidana yang dijatuhkan; <u>Menetapkan</u> agar <u>Terdakwa</u> tetap berada dalam tahanan; <u>Menetapkan</u> barang bukti berupa: 1 (satu) buah <u>HP I Phone 6</u> warna gold dengan <u>No. Simcard 0822 2172 2002</u> yang masih diperlukan sebagai barang bukti dalam perkara <u>No 509/Pid.Sus/2019/PN Smn</u>, maka dikembalikan kepada <u>Penuntut Umum</u> untuk dijadikan barang bukti dalam perkara <u>No 509/Pid.Sus/2019/PN Smn</u>; <u>Membebaskan</u> kepada <u>Terdakwa</u></p>	<p><u>mengadili</u>: <u>menyatakan</u> <u>terdakwa pulung aji wibowo bin hudi sumarno alm</u> telah terbukti secara sah dan meyakinkan bersalah melakukan tindak pidana bersama-sama melakukan penyalahgunaan narkotika golongan <u>i</u> bagi diri sendiri; menjatuhkan pidana kepada <u>terdakwa</u> oleh karena itu dengan pidana penjara selama 8 (delapan) bulan; <u>menetapkan</u> masa penangkapan dan penahanan yang telah dijalani <u>terdakwa</u> dikurangkan seluruhnya dari pidana yang dijatuhkan; <u>menetapkan</u> agar <u>terdakwa</u> tetap berada dalam tahanan; <u>menetapkan</u> barang bukti berupa: 1 (satu) buah <u>hp i phone 6</u> warna gold dengan <u>no. simcard 0822 2172 2002</u> yang masih diperlukan sebagai barang bukti dalam perkara <u>no 509/pid.sus/2019/pn smn</u>, maka dikembalikan kepada <u>penuntut umum</u> untuk dijadikan barang bukti dalam perkara <u>no 509/pid.sus/2019/pn smn</u>; <u>membebaskan</u> kepada <u>terdakwa</u> membayar biaya perkara sejumlah <u>rp 2000</u> (dua ribu rupiah);</p>

Sebelum	Sesudah
membayar biaya perkara sejumlah Rp 2000 (dua ribu rupiah);	

b. *Remove Punctuation*

Remove punctuation digunakan untuk menghapus karakter tanda baca seperti tanda tanya, koma, titik dua, dan lain sebagainya. Hasil dari *remove punctuation* dapat dilihat pada tabel 5.3 berikut.

Tabel 5.3 Hasil *Remove Punctuation*

Sebelum	Sesudah
mengadili; menyatakan terdakwa pulung aji wibowo bin hudi sumarno alm telah terbukti secara sah dan meyakinkan bersalah melakukan tindak pidana bersama-sama melakukan penyalahgunaan narkotika golongan i bagi diri sendiri; menjatuhkan pidana kepada terdakwa oleh karena itu dengan pidana penjara selama 8 (delapan) bulan; menetapkan masa penangkapan dan penahanan yang telah dijalani terdakwa dikurangkan seluruhnya dari pidana yang dijatuhkan; menetapkan agar terdakwa tetap berada dalam tahanan; menetapkan barang bukti berupa: 1 (satu) buah hp i phone 6 warna gold dengan no. simcard 0822 2172 2002 yang masih diperlukan sebagai barang bukti dalam perkara no	mengadili menyatakan terdakwa pulung aji wibowo bin hudi sumarno alm telah terbukti secara sah dan meyakinkan bersalah melakukan tindak pidana bersama-sama melakukan penyalahgunaan narkotika golongan i bagi diri sendiri menjatuhkan pidana kepada terdakwa oleh karena itu dengan pidana penjara selama 8 delapan bulan menetapkan masa penangkapan dan penahanan yang telah dijalani terdakwa dikurangkan seluruhnya dari pidana yang dijatuhkan menetapkan agar terdakwa tetap berada dalam tahanan menetapkan barang bukti berupa 1 satu buah hp i phone 6 warna gold dengan no simcard 0822 2172 2002 yang masih diperlukan sebagai barang bukti dalam perkara no

Sebelum	Sesudah
509/pid_sus/2019/pn smn, maka dikembalikan kepada penuntut umum untuk dijadikan barang bukti dalam perkara no 509/pid_sus/2019/pn smn; membebankan kepada terdakwa membayar biaya perkara sejumlah rp 2000 (dua ribu rupiah);	509pidsus2019pn smn maka dikembalikan kepada penuntut umum untuk dijadikan barang bukti dalam perkara no 509pidsus2019pn smn membebankan kepada terdakwa membayar biaya perkara sejumlah rp 2000 dua ribu rupiah

c. *Stopword removal*

Stopword removal digunakan untuk penghapusan kata yang sering muncul namun tidak memiliki makna. Contoh kata yang termasuk dalam *stopword removal* dalam Bahasa Indonesia adalah “dan”, “atau”, “yang”, “dengan”, dan lain sebagainya. Hasil dari *stopword removal* dapat dilihat pada tabel 5.4 berikut.

Tabel 5.4 Hasil *Stopword Removal*

Sebelum	Sesudah
<u>mengadili</u> <u>menyatakan</u> <u>terdakwa</u> pulung aji wibowo <u>bin</u> hudi sumarno <u>alm</u> <u>telah</u> terbukti <u>secara</u> sah <u>dan</u> meyakinkan bersalah <u>melakukan</u> <u>tindak</u> <u>pidana</u> <u>bersama-sama</u> <u>melakukan</u> penyalahgunaan narkotika golongan i <u>bagi</u> diri sendiri menjatuhkan <u>pidana kepada terdakwa</u> <u>oleh karena itu dengan pidana</u> penjara <u>selama</u> 8 delapan bulan menetapkan <u>masa</u> penangkapan <u>dan</u> penahanan <u>yang telah</u> dijalani <u>terdakwa</u> dikurangkan <u>seluruhnya dari pidana</u> <u>yang</u> dijatuhkan menetapkan <u>agar</u>	pulung aji wibowo hudi sumarno terbukti sah meyakinkan bersalah penyalahgunaan narkotika golongan i diri sendiri menjatuhkan penjara 8 delapan bulan menetapkan penangkapan penahanan dijalani dikurangkan dijatuhkan menetapkan tahanan menetapkan barang bukti 1 buah hp i phone 6 warna gold no simcard 0822 2172 2002 diperlukan barang bukti no 509pidsus2019pn smn dikembalikan penuntut umum dijadikan barang bukti no

Sebelum	Sesudah
<p><u>terdakwa tetap berada dalam</u> tahanan menetapkan barang bukti <u>berupa</u> 1 <u>satu</u> buah hp i phone 6 warna gold <u>dengan</u> no simcard 0822 2172 2002 <u>yang masih</u> diperlukan <u>sebagai</u> barang bukti <u>dalam</u> <u>perkara</u> no 509pidsus2019pn smn <u>maka</u> dikembalikan <u>kepada</u> penuntut umum <u>untuk</u> dijadikan barang bukti <u>dalam</u> <u>perkara</u> no 509pidsus2019pn smn membebankan <u>kepada</u> <u>terdakwa</u> membayar biaya <u>perkara sejumlah</u> rp 2000 <u>dua</u> ribu rupiah</p>	<p>509pidsus2019pn smn membebankan membayar biaya rp 2000 ribu rupiah</p>

d. *Tokenizing*

Tokenizing digunakan untuk memisahkan daretan kata sehingga diperoleh potongan kata atau token yang akan menjadi entitas yang memiliki nilai dalam penyusunan matriks dokumen pada proses selanjutnya. Visualisasi dari *tokenizing* dapat dilihat pada tabel 5.5 berikut.

Tabel 5.5 Hasil *Tokenizing*

Sebelum	Sesudah
<p>pulung aji wibowo hudi sumarno terbukti sah meyakinkan bersalah penyalahgunaan narkoba golongan i diri sendiri menjatuhkan penjara 8 delapan bulan menetapkan penangkapan penahanan dijalani dikurangkan dijatuhkan menetapkan tahanan menetapkan barang bukti 1</p>	<p>['pulung', 'aji', 'wibowo', 'hudi', 'sumarno', 'terbukti', 'sah', 'meyakinkan', 'bersalah', 'penyalahgunaan', 'narkoba', 'golongan', 'i', 'diri', 'sendiri', 'menjatuhkan', 'penjara', '8', 'delapan', 'bulan', 'menetapkan', 'penangkapan', 'penahanan',</p>

Sebelum	Sesudah
buah hp i phone 6 warna gold no simcard 0822 2172 2002 diperlukan barang bukti no 509pidsus2019pn smn dikembalikan penuntut umum dijadikan barang bukti no 509pidsus2019pn smn membebankan membayar biaya rp 2000 ribu rupiah	‘dijalani’, ‘dikurangkan’, ‘dijatuhkan’, ‘menetapkan’, ‘tahanan’, ‘menetapkan’, ‘barang’, ‘bukti’, ‘1’, ‘buah’, ‘hp’, ‘i’, ‘phone’, ‘6’, ‘warna’, ‘gold’, ‘no’, ‘simcard’, ‘082221722002’, ‘diperlukan’, ‘barang’, ‘bukti’, ‘no’, ‘509pidsus2019pn’, ‘smn’, ‘dikembalikan’, ‘penuntut’, ‘umum’, ‘dijadikan’, ‘barang’, ‘bukti’, ‘no’, ‘509pidsus2019pn’, ‘smn’, ‘membebankan’, ‘membayar’, ‘biaya’, ‘rp’, ‘2000’, ‘ribu’, ‘rupiah’]

Setelah dilakukan preprocessing data kemudian dilanjutkan ke tahap pembobotan kata menggunakan TF-IDF.

5.3 Pembobotan Kata

Dalam melakukan analisis *topic modeling*, biasanya data yang digunakan harus berbentuk numerik. Sehingga untuk mengubah data tersebut menjadi data numeric digunakan metode pembobotan TF-IDF. Dalam penulisan ini, penulis mengambil beberapa sampel kata yang ada pada dokumen seperti kata ahli, alkohol, apotik, atm, dan zinah. Berikut merupakan tabel hasil pembobotan kata TF-IDF dari sampel kata yang penulis gunakan.

Tabel 5.6 Tabel Sampel Hasil Pembobotan Kata TF-IDF

<i>D</i>	Ahli	Alkohol	Apotik	ATM	...	Zinah
	∴	∴	∴	∴	∴	∴
607	8,695303	0	0	0	0	0
668	0		8,695303	0	0	0
819	0	0	0	0	0	6,615862

D	Ahli	Alkohol	Apotik	ATM	...	Zinah
1567	0	8,695303	0	0	0	0
1760	0	0	0	8,695303	0	0

Berdasarkan tabel 5.6 di atas, tabel tersebut merupakan contoh kata yang menghasilkan nilai TF-IDF dihitung secara otomatis menggunakan *software Phyton*. Untuk perhitungan manualnya, pertama mengitung *Term Frequency* (TF) terlebih dahulu. Berikut merupakan tabel hasil TF pada sampel kata yang penulis gunakan.

Tabel 5.7 Tabel sampel hasil nilai TF

D	Ahli	Alkohol	Apotik	ATM	...	Zinah
	∴	∴	∴	∴	∴	∴
607	1	0	0	0	0	0
668	0	0	1	0	0	0
819	0	0	0	0	0	1
1567	0	1	0	0	0	0
1760	0	0	0	1	0	0

Pada perhitungan manual, sebagai contoh kata “alkohol” yang berada pada dokumen ke 1567. Pada tabel 5.7 merupakan hasil dari perhitungan TF secara manual. Nilai 0 pada tabel TF berarti bahwa dalam 1 dokumen kata yang dicari tidak ada dalam dokumen tersebut, sedangkan nilai 1 pada tabel TF yang berarti bahwa dalam 1 dokumen kata yang dicari muncul satu kali, dan seterusnya. Selanjutnya menghitung nilai DF secara manual yaitu jumlah dokumen pada suatu term yang memunculkan kata “alkohol”. Nilai DF dari kata “alkohol” dari semua dokumen sebanyak 1. Kemudian dihitung nilai IDF atau *inverse* dari DF, *term* yang memiliki DF yang rendah memiliki nilai IDF yang tinggi. Perhitungan nilai IDF dapat menggunakan rumus seperti pada persamaan 3.3 di atas. Dalam penelitian ini menggunakan jumlah dokumen sebanyak 2198 dokumen. Berikut merupakan perhitungan nilai IDF.

$$IDF = \ln\left(\frac{DF}{DF_t}\right)$$

$$IDF = \ln\left(\frac{2198}{1}\right) = 7,695303$$

Setelah mendapatkan nilai TF, DF, dan IDF, selanjutnya menghitung nilai pembobotan TF-IDF secara manual, yaitu dengan mengalikan nilai TF dengan nilai IDF dengan menggunakan persamaan 3.5. Berikut merupakan perhitungan pembobotan TF-IDF.

$$w = TF \times (IDF + 1)$$

$$w = 1 \times 7,695303 = 8,695303$$

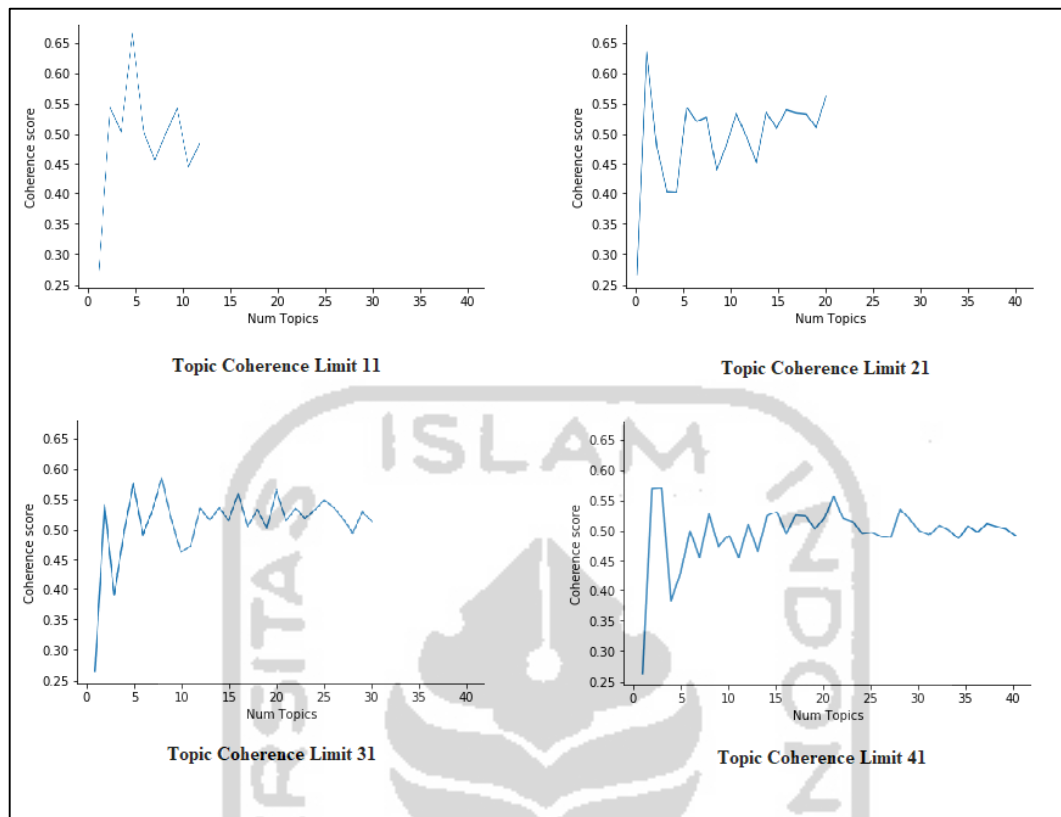
Dari hasil pembobotan di atas maka dapat disimpulkan bahwa kata “alkohol” pada dokumen ke 1567 mempunyai pembobotan kata TF-IDF sebesar 8,695303.

Setelah dilakukan pembobotan kata secara manual, kemudian hasil pembobotan kata secara manual dibandingkan dengan hasil otomatis menggunakan *software Phyton* pada tabel 5.6. dapat diketahui hasil dari pembobotan yang telah dilakukan sam dengan hasil otomatis menggunakan *software*. Hasil lengkap pembobotan TF-IDF manual dan *software* terdapat dilampiran.

Langkah selanjutnya setelah melakukan pembobotan kata menggunakan TF-IDF yaitu melakukan analisis *topic modeling* dengan menggunakan LDA.

5.4 Topic Modeling dengan Latent Dirichlet Allocation (LDA)

Menentukan jumlah model topik dalam *topic modeling* dapat dilakukan dengan melihat visualisasi pada gambar grafik *topics coherence*. Semakin tinggi nilai *topics coherence* menunjukkan bahwa model yang dihasilkan akan semakin baik. Sehingga jumlah topik dengan nilai *topics coherence* tertinggi yang akan digunakan dalam *topic modeling*.



Gambar 5.3. Grafik *Coherence Score*

Gambar 5.3 di atas menunjukkan visualisasi grafik *topics coherence*. Pada penelitian ini, peneliti menggunakan perbandingan nilai limit 11, limit 21, limit 31, dan limit 41. Penggunaan nilai limit sebenarnya tergantung dengan peneliti itu tersendiri, disini peneliti menggunakan nilai-nilai limit tersebut karena peneliti mengacu pada penelitian (Listari, 2019).

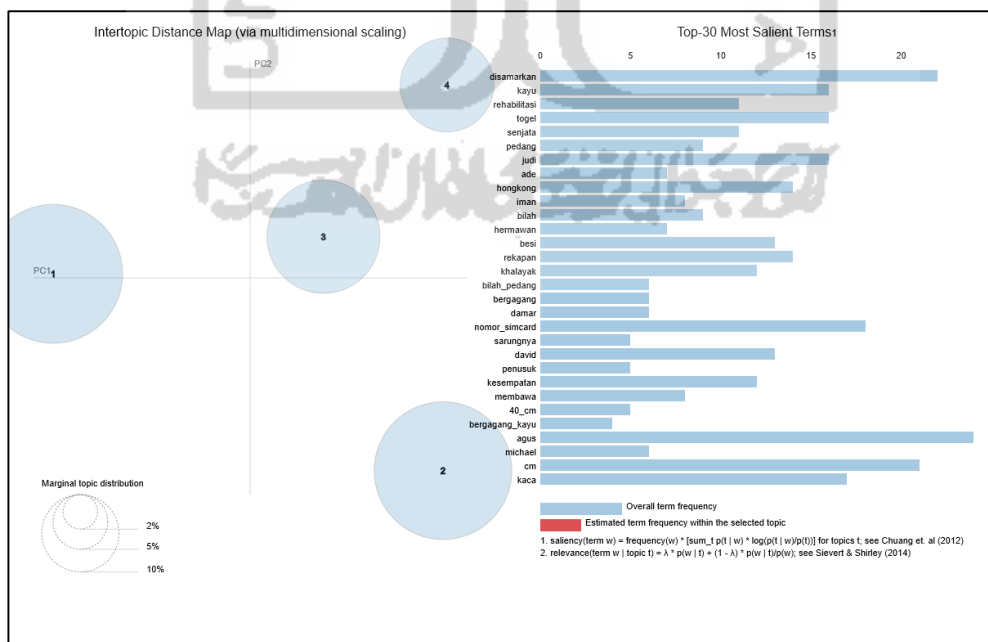
Dari gambar 5.3 di atas dapat dilihat bahwa pada *topic coherence limit 11* dengan *num topics* ke 4 yang grafik atau nilainya paling tinggi, sehingga pada penulisan ini menggunakan 4 topik pada *topic coherence limit 11*. Namun, gambar 5.3 tersebut hanya menunjukkan visualisasinya saja. Dalam memudahkan melihat visualisasi pada gambar grafik *topics coherence* maka dapat dilihat nilai *topic coherence limit 11* pada tabel 5.8 berikut, sedangkan untuk nilai *topic coherence* lainnya dapat dilihat pada lampiran 3.

Tabel 5.8 Nilai *Topic Coherence* limit 11

<i>Num Topics</i>	<i>Coherence Value</i>	<i>Num Topics</i>	<i>Coherence Value</i>
1	0,26257	6	0,45086
2	0,53723	7	0,49529
3	0,49773	8	0,53551
4	0,65895	9	0,44025
5	0,49529	10	0,47719

Hasil *topics coherence* pada tabel 5.8 dan lampiran 3 menunjukkan bahwa nilai dari *topics coherence* yang paling tinggi yaitu pada *num topics* ke-4 pada *topic coherence limit 11* dengan nilai *coherence value*-nya sebesar 0,65895. Berdasarkan nilai *coherence value* terbaik, maka jumlah topik yang dihasilkan tersebut akan dijadikan acuan dalam pembuatan model, sehingga dalam penelitian ini digunakan empat model topik yang terbaik. Hasil ini sama dengan hasil visualisasi pada gambar 5.3 di atas.

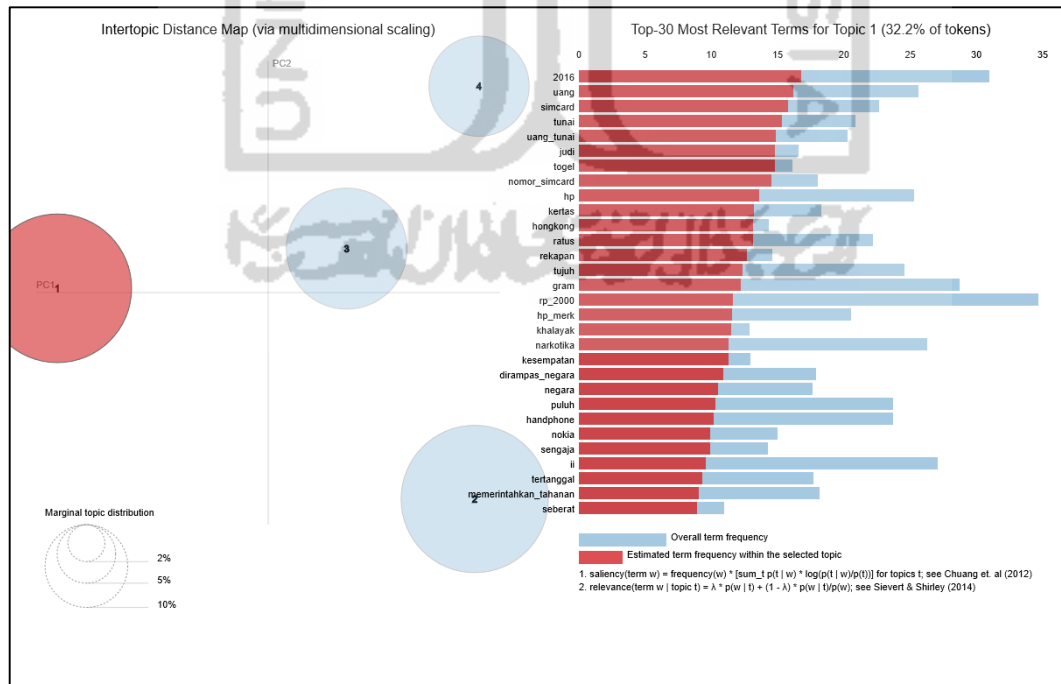
Selanjutnya dilakukan analisis *topic modeling* dengan melihat visualisasi dari *pyLDAvis* dan keterkaitan antar kata yang dihasilkan, kemudian disimpulkan dalam suatu topik tertentu. Gambar 5.4 berikut merupakan visualisasi dari *pyLDAvis* dengan dua topik yang diperoleh dari hasil *topics coherence*.



Gambar 5.4. Visualisasi Topic Modeling dengan *PyLDAvis*

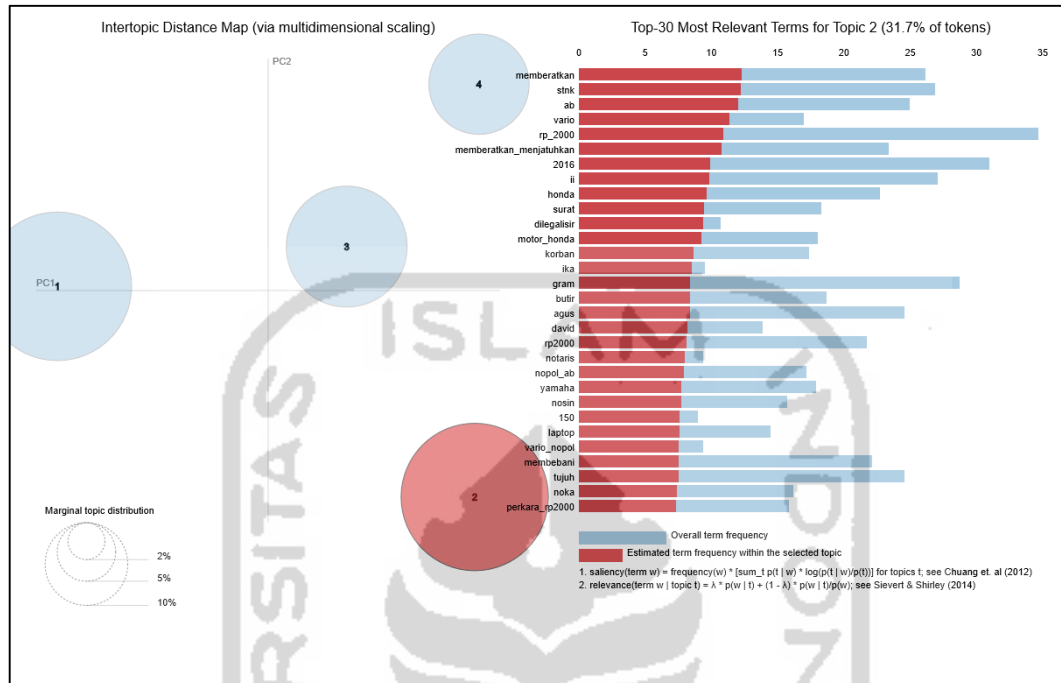
Berdasarkan gambar 5.4 pada bagian kiri memperlihatkan representasi topik secara global dimana dibuat pemetaan jarak antar topik (*intertopic distance map*) melalui multidimensional scaling dengan jumlah kluster topik sebanyak 2 topik. Besar kecilnya suatu kluster topik menjelaskan tentang ukuran seberapa pentingnya topik terhadap kata, dengan posisi dokumen dijelaskan oleh wilayah *latent* yang terdeteksi selama inferensi. Pada visualisasi LDA dibagi menjadi 2 bagian sumbu *Principal Component* yaitu PC1 dan PC2 sehingga terbentuk kuadran-kuadran dimana lingkaran-lingkaran topik yang berdekatan memiliki hubungan yang kuat. Pada lingkaran yang saling berdekatan tersebut artian kata-kata yang ada cenderung sama. Sehingga dapat dikatakan semakin berdekatan suatu lingkaran topik dengan yang lainnya maka artian kata-kata yang ada semakin cenderung sama. Model topik yang bagus memiliki artian kata-kata yang berbeda dengan topik yang lainnya.

Pada bagian kanan gambar 5.4 memperlihatkan *bar chart top 30* global terminologi paling relevan pada suatu topik tertentu. *Bar chart* yang berwarna biru yang mengindikasikan keseluruhan *term frequency* dalam *corpus*, sedangkan *bar chart* yang berwarna merah menunjukkan estimasi *term frequency* terhadap suatu topik tertentu.



Gambar 5.5. Visualisasi Topik 1

penulis dapat menyimpulkan bahwa bahasan pada topik 1 cenderung membahas mengenai kasus perjudian.



Gambar 5.8. Visualisasi Topik 2

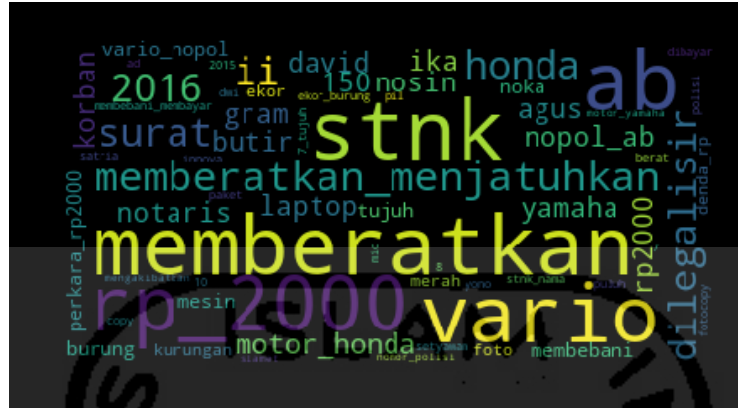
Gambar 5.8 di atas menunjukkan topik 2 yaitu lingkaran yang berwarna merah. Marginal topic distribution pada topik 2 adalah 31,7% yang berarti bahwa pentingnya topik 2 terhadap corpus sebesar 31,7%. Pada sisi kanan terdapat bar chart dengan warna merah memperlihatkan 30 terminologi paling relevan terhadap topik 2 tersebut, sedangkan yang berwarna biru menunjukkan keseluruhan dari term frequency pada corpus. Model pada topik 2 ini dapat dilihat pada gambar 5.9 berikut.

```
[(0,
'0.003*"memberatkan" + 0.003*"stnk" + 0.003*"ab" + 0.003*"vario" + '
'0.003*"rp_2000" + 0.003*"memberatkan_menjatuhkan" + 0.003*"2016" + '
'0.003*"ii" + 0.003*"honda" + 0.003*"surat")],
```

Gambar 5.9. Model Topik 2

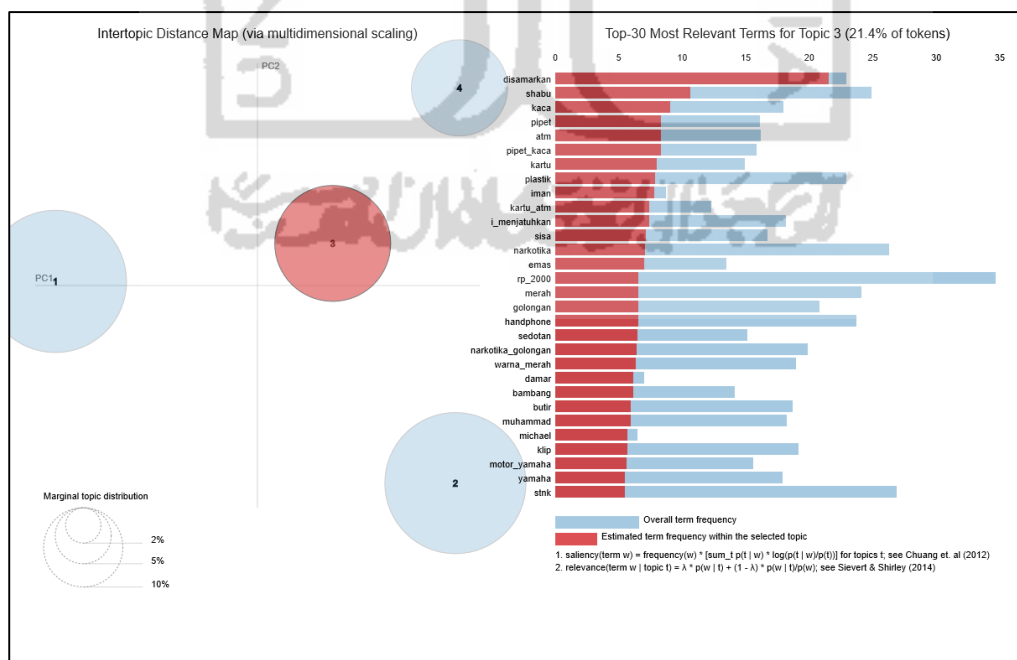
Berdasarkan gambar 5.9 di atas model pada topik 2, lima term yang memiliki peluang penyebutan paling besar pada topik 2 ini adalah kata disamarkan dengan probabilitas paling besar yaitu 0,009 kemudian diikuti dengan kata shabu dengan probabilitas 0,004; kata kaca dengan probabilitas 0,004; kata pipet dengan

probabilitas 0,003, dan kata atm dengan probabilitas 0,003. Berikut ini merupakan word cloud dari topik 2.



Gambar 5.10. Word Cloud Topik 2

Dapat dilihat gambar 5.10 di atas menunjukkan *Word cloud* paling relevan topik 2. Word cloud tersebut merepresentasikan dari 30 terminologi paling relevan terhadap topik 2. Ketika probabilitas pada suatu term lebih besar dibandingkan dengan yang lain maka visualisasi term tersebut word cloud akan lebih menonjol atau lebih besar. Berdasarkan 30 terminologi paling relevan terhadap topik 2 maka penulis dapat menyimpulkan bahwa bahasan pada topik 2 cenderung membahas mengenai kasus pencurian.



Gambar 5.11. Visualisasi Topik 3

Gambar 5.11 di atas menunjukkan topik 3 yaitu lingkaran yang berwarna merah. Marginal topic distribution pada topik 3 adalah 21,4% yang berarti bahwa pentingnya topik 3 terhadap corpus sebesar 21,4%. Pada sisi kanan terdapat bar chart dengan warna merah memperlihatkan 30 terminologi paling relevan terhadap topik 3 tersebut, sedangkan yang berwarna biru menunjukkan keseluruhan dari term frequency pada corpus. Model pada topik 2 ini dapat dilihat pada gambar 5.12 berikut.

```
(1,
'0.009*"disamarkan" + 0.004*"shabu" + 0.004*"kaca" + 0.003*"pipet" + '
'0.003*"atm" + 0.003*"pipet_kaca" + 0.003*"kartu" + 0.003*"plastik" + '
'0.003*"iman" + 0.003*"kartu_atm"'),
```

Gambar 5.12. Model Topik 3

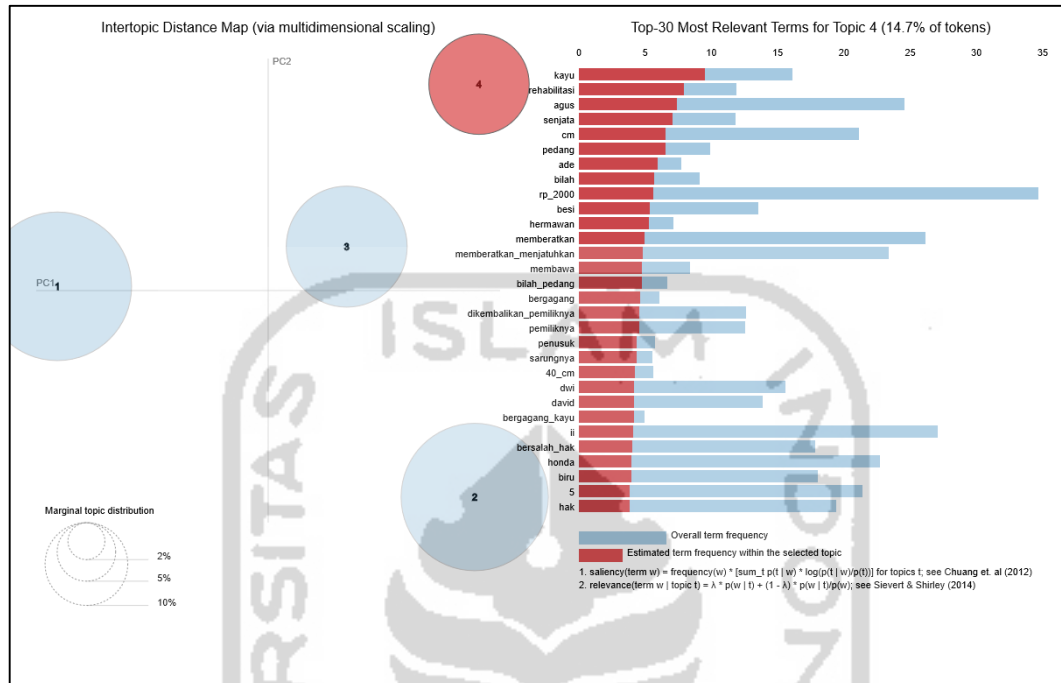
Berdasarkan gambar 5.12 di atas model pada topik 3, lima term yang memiliki peluang penyebutan paling besar pada topik 3 ini adalah kata kayu dengan probabilitas paling besar yaitu 0,006 kemudian diikuti dengan kata rehabilitas dengan probabilitas 0,005; kata agus dengan probabilitas 0,004; kata senjata dengan probabilitas 0,004, dan kata cm dengan probabilitas 0,004. Berikut ini merupakan word could dari topik 3.



Gambar 5.13. Word Could Topik 3

Dapat dilihat gambar 5.13 di atas menunjukkan *Word could* paling relevan topik 3. Word could tersebut merepresentasikan dari 30 terminologi paling relevan terhadap topik 3. Ketika probabilitas pada suatu term lebih besar dibandingkan dengan yang lain maka visualisasi term tersebut word could akan lebih menonjol atau lebih besar. Berdasarkan 30 terminologi paling relevan terhadap topik 3 maka

penulis dapat menyimpulkan bahwa bahasan pada topik 3 cenderung membahas mengenai kasus narkoba.



Gambar 5.14. Visualisasi Topik 4

Gambar 5.14 di atas menunjukkan topik 4 yaitu lingkaran yang berwarna merah. Marginal topic distribution pada topik 4 adalah 14,7% yang berarti bahwa pentingnya topik 4 terhadap corpus sebesar 14,7%. Pada sisi kanan terdapat bar chart dengan warna merah memperlihatkan 30 terminologi paling relevan terhadap topik 2 tersebut, sedangkan yang berwarna biru menunjukkan keseluruhan dari term frequency pada corpus. Model pada topik 4 ini dapat dilihat pada gambar 5.15 berikut.

$$(2, '0.006**"kayu" + 0.005**"rehabilitasi" + 0.004**"agus" + 0.004**"senjata" + '0.004**"cm" + 0.004**"pedang" + 0.003**"ade" + 0.003**"bilah" + 0.003**"rp_2000" ' + 0.003**"besi"'),$$

Gambar 5.15. Model Topik 4

Berdasarkan gambar 5.15 di atas model pada topik 4, lima term yang memiliki peluang penyebutan paling besar pada topik 4 ini adalah term 2016 dengan probabilitas paling besar yaitu 0,004 kemudian diikuti dengan kata uang dengan probabilitas 0,004; kata simcard dengan probabilitas 0,004; kata tunai

pencurian. Hal tersebut berkaitan dengan hasil yang diperoleh menurut Badan Pusat Statistik (BPS) Yogyakarta pada tahun 2018 adalah kejahatan pencurian dengan persentase sebanyak 67,12 persen.



BAB 6 PENUTUP

6.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan dalam penelitian pada penulisan ini, maka dapat diambil kesimpulan sebagai berikut.

1. Gambaran umum data perkara di Pengadilan Negeri Sleman pada tahun 2016 – Januari 2020, yaitu 5 perkara yang paling menonjol, diantaranya pencurian, narkoba, penggelapan, penipuan, dan perjudian. Angka perkara yang paling tinggi adalah pencurian dengan jumlah kasus sebanyak 572 kasus. Sedangkan jumlah orang yang melakukan tindak pidana didominasi oleh kaum pria dengan persentase sebanyak 90%.
2. Dalam menggunakan *topic modeling* dengan metode *Latent Dirichlet Allocation* (LDA) jumlah topik yang terbentuk terdapat empat topik. Topik 1 cenderung membahas kasus perjudian, topik 2 cenderung membahas kasus pencurian/penggelapan/penipuan, topik 3 cenderung membahas kasus narkoba, sedangkan topik 4 cenderung membahas kasus senjata api dan benda tajam / kekerasan/ penganiayaan.

6.2 Saran

Dari hasil penelitian ini, maka peneliti memberikan beberapa saran sebagai berikut.

1. Pada penelitian ini, peneliti hanya menggunakan data perkara pada tahun 2016 – Januari 2020. Untuk penelitian selanjutnya, sebaiknya menambahkan data perkara sehingga hasil yang didapatkan lebih baik dan spesifik.
2. Pada penelitian ini, peneliti hanya menggunakan analisis *topic modeling* dengan metode LDA. Untuk penelitian selanjutnya diharapkan dapat mengembangkan analisis yang akan digunakan contohnya dengan mengklasifikasikan tempat terjadinya perkara sehingga dapat diketahui di daerah mana perkara tersebut sering terjadi.

3. Untuk aparat penegak hukum khususnya di daerah Kabupaten Sleman, D.I. Yogyakarta lebih menegakkan hukum dalam menangani tindak pidana khususnya pada pidana yang sering terjadi di daerah Kabupaten Sleman karena angka kriminalitas di Kabupaten Sleman paling tinggi dibandingkan dengan kabupaten/kota lainnya di D.I. Yogyakarta.
4. Untuk admin pada *website* <http://pn-sleman.go.id/> disarankan untuk mengembangkan mesin *topic modeling* untuk informasi secara otomatis sehingga ketika terdapat putusan baru maka langsung memperbarui topik, hal ini dapat dijadikan sebagai bahan pembelajaran untuk masyarakat umum terkait putusan hakim pada tindak pidana.
5. Untuk khalayak umum disarankan untuk lebih meningkatkan kewaspadaan terhadap diri sendiri.



DAFTAR PUSTAKA

- Ahmad, A. (2017). Mengenal Artificial Intelligence, Machine Learning, Neural Network, dan Deep Learning. *Jurnal Teknologi Indonesia*, 1-5.
- Andayu, N. P. (2013). Perancangan Text To Speech Converter Engine dalam Pengucapan Kata Berbahasa Arab Sehari-hari. *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, 1-6.
- Astuti, R. (2014). Pengolahan Data Statistik Hasil Pembangunan Infrastruktur Pekerjaan Umum. *Media Informatika*, 33-47.
- Blei, D. M. (2012). Probabilistic Topic Models. *Communication of the ACM*, 77-84.
- Blei, D. M. (2013). Topic Modeling and Digital Humanities. *Journal of Digital Humanities*.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 993-1022.
- Brett, M. R. (2012, December 12). *Topic Modeling: A Basic Introduction*. Retrieved from Journal of Digital Humanities: <http://journalofdigitalhumanities.org/2-1/topic-modeling-a-basic-introduction-by-megan-r-brett/>
- Effendi, T. (2016). *Praktik Peradilan Pidana*. Malang: Setara Press.
- ELvira, J. B., Cobo, M., Viedma, E. H., & Herrera, A. L. (2019). Latent Dirichlet Allocation (LDA) for Improving The Topic Modeling of The Official Bulletin of The Spanish State (BOE). *Procedia Computer Science*, 207-214.
- Fajriyanto, M. (2018). Penerapan Metode Bayesian dalam Model Latent Dirichlet Allocation Di Media Sosial. *Jurnal Pendidikan Matematika dan Sains*, 1-6.
- Fitriasih, M., & Kusumaningrum, R. (2019). ANALISIS KLASIFIKASI OPINI TWEET PADA MEDIA SOSIAL TWITTER MENGGUNAKAN LATENT DIRICHLET ALLOCATION (LDA). *Seminar Nasional Teknologi Informasi dan Komunikasi*, 177-186.

- Hariyady, Basuki, S., & Meidina, L. (2018). Implementasi Algoritma Deteksi Topik Keluhan Pelanggan Jasa Ojek Online Berdasarkan Komentar Media Sosial. *Seminar Nasional Teknologi dan Rekayasa*, 160-166.
- Hartanti, E. (2009). *Tindak Pidana Korupsi*. Jakarta: Sinar Grafika.
- Hasanah, S. (2017, Juli 24). *Perbedaan Tindak Pidana Ringan (Tipiring) dengan Pelanggaran*. Retrieved Mei 15, 2020, from Hukum Online: <https://www.hukumonline.com/klinik/detail/ulasan/lt5971008e81638/perbedaan-tindak-pidana-ringan-tipiring-dengan-pelanggaran/>
- Hikmah, N. (2018). Pemanfaatan Text Mining dalam Pencarian Ayat Al Quran menggunakan TF-IDF dan Cosine Similarity. *Antartika*, 8, 1-61.
- Ilyas, A. (2012). *ASAS-ASA HUKUM PIDANA MEMAHAMI TINDAK PIDANA DAN PERTANGGUNGJAWABAN PIDANA SEBAGAI SYARAT PEMIDANAAN*. Yogyakarta: Rangkang Education Yogyakarta & PuKAP-Indonesia.
- Jing, Q. (2014). *Searching for economic effects of user specified events based on topic modelling and event reference*. Retrieved from Acadia Scholar: <https://scholar.acadiau.ca/islandora/object/theses%3A341>
- Keamanan, S. P. (2014). *Statistik Kriminal 2014*. Jakarta: Badan Pusat Statistik.
- Listari. (2019). *INISIASI NATURAL LANGUAGE PROCESSING (NLP) DAN KLASIFIKASI JENIS WISATA KULINER UNTUK PROGRAM CHATBOT*. Yogyakarta: Universitas Islam Indonesia.
- Maulana, G. Y. (2017). *Aplikasi Self Organizing Maps (SOM) untuk Pengelompokan Tindak Pidana Di Kabupaten Sleman Daerah Istimewa Yogyakarta*. Yogyakarta: Universitas Islam Indonesia.
- Mursalin, I. A. (2012, September 2). *Tindakan Kriminal Dan Kejahatan*. Retrieved April 11, 2020, from scribd: <https://www.scribd.com/doc/104704488/Tindakan-Kriminal-Dan-Kejahatan>
- Ningsah, K. R., & Kuncoro, J. (2017). Persepsi Terhadap Perilaku Tindak Kriminal Ditinjau Dari Kepribadian The Big Five & Status Hukum Wanita Narapidana & Wanita Non Narapidana. *Proyeksi*, 27-34.

- Nurjannah, M., Hamdani, & Astuti, I. F. (2013). Penerapan Algoritma Term Frequency-Inverse Document Frequency (TF-IDF) untuk Text Mining. *Informatika Mulawarman*, 8, 111.
- Porter, K. (2018). Analyzing the DarkNetMarkets Subreddit for Evolutions of Tools and Trends Using LDA Topic Modeling. *Digital Investigation*, 87-97.
- Prasetyo, T. (2015). *Hukum Pidana*. Jakarta: PT RajaGrafindo Persada.
- Putra, I. M. (2017). Analisis topik Informasi Publik Media Sosial Di Surabaya menggunakan Pemodelan Latent Dirichlet Allocation (LDA).
- Putri, R. C., & Susilo, B. (2019). Pola Spasional Pencurian Kendaraan Bermotor (Curanmor) Di Daerah Istimewa Yogyakarta Tahun 2018. *Jurnal Bumi Indonesia*, 1-6.
- Rauhan, A. (2019). Pengolahan Data Menggunakan Machine Learning. *Repository Pertamina University*, 11-15.
- Setiady, T. (2010). *Pokok-Pokok Penitensier Indonesia*. Bandung: Alfabeta.
- Siregar, R. R., Sinaga, F. A., & Arianto, R. (2017). Aplikasi Penentuan Dosan Penguji Skripsi Menggunakan Metode TF-IDF dan Vector Space Model. *Journal of Computer and information Systems*, 1, 171-186.
- Sleman, P. N. (2015). *Tugas Pokok dan Fungsi*. Retrieved Februari 14, 2020, from Pengadilan Negeri Sleman: <http://pn-sleman.go.id/new//link/2016060115153026208574e8b9258003.html>
- Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012). Exploring Topic Coherence Over Many Models and Many Topics. *Empirical Methods in Natural language Processing and Computational Natural Language Learning* (pp. 952-961). Jeju Island, Korea: Association for Computational Linguistics.
- Tang, F. (2019, July 14). *Beginner's Guide to LDA Topic Modeling with R*. Retrieved Maret 20, 2020, from Toward Data Science: <https://towardsdatascience.com/beginners-guide-to-lda-topic-modelling-with-r-e57a5a8e7a25>
- Tomalili, R. (2015). *Hukum Pidana*. Yogyakarta: Deepublish.

- Tong, Z., & Zhang, H. (2016). A Text Mining Research Basen on LDA Topic Modeling. *Jodrey School of Computer Science, Acadia University, Wolfville, NS, Canada, 10.5121*, 201-210.
- Usfa, A. F., & Tongat. (2004). *Pengantar Hukum Pidana*. Malang: UMM Press.
- Wibowo, A. (2019). *Statistik Politik dan Keamanan Daerah Istimewa Yogyakarta 2018*. Yogyakarta: Badan Pusat Statistik Provinsi Daerah Istimewa Yogyakarta.
- Wijaya, E. (2013). Analisis Penggunaan Algoritma Breadth First Search Dalam Konsep Artificial Intelligence. *Jurnal TIME*, 2, 18-26.



Perundang-Undangan

Pasal 10 KUHP.

Pasal 18 KUHP.

Pasal 35 Ayat 1 KUHP.

Pasal 38 Ayat 1 KUHP.

Pasal 39 KHUP.

Pasal 69 Ayat 1 KUHP.

UU No. 20 Tahun 1946 Pasal 1.

UU No. 20 Tahun 1946 Pasal 2.

UU No. 8 Tahun 1981 KUHP.

Pasal 152-202 KUHAP.

Pasal 203-204 KUHAP.

Pasal 205-216 KUHAP.



LAMPIRAN



Lampiran 1 Data perkara Pengadilan Negeri Sleman tahun 2016 – Januari 2020

No	Nomor Pe Tgl	Registras	Klasifikasi Perkara	Penuntut Terdakwa	Status Per	Lama Pro	Jenis Perk	Dakwaan	Pihak Dipi	Penuntut/ Putusan S	Putusan	Klasifikasi singkat	Lama Kurungan
1	619/Pid.B	21-Dec-16	Kejahatan Perjudi	Andy Nug Walijo; Ferdianto	Minutasi	35	Biasa	bahwa Ter Ya		Menuntut Menyatak	Mengadili	lv perjudian	4 bulan ; 4 bulan
2	620/Pid.B	21-Dec-16	Kejahatan Perjudi	Andy Nug widiyono bin ketoarj	Minutasi	48	Biasa	Bahwa ter Ya		Menuntut Supaya M	Mengadili	lv perjudian	1 bulan ; 1 bulan
3	614/Pid.B	20-Dec-16	Penganiayaan	Siti Mahm Rahmat hidayat alias	Minutasi	42	Biasa	bahwa ter Ya		Menuntut Menyatak	Mengadili	lv kekerasan	1 tahun
4	615/Pid.B	20-Dec-16	Penggelapan	Dhudi Hac rendy cahyo nugroho	Minutasi	62	Biasa	mengadili Ya		Menuntut Supaya M	Mengadili	lv penggelapan	1 tahun
5	616/Pid.B	20-Dec-16	Pencurian	Handry Su purwanta alias pur bir	Minutasi	41	Biasa	Bahwa ter Ya		Menuntut Supaya M	Mengadili	lv pencurian	3 bulan
6	617/Pid.B	20-Dec-16	Penipuan	Christina festher novianty binti	Minutasi	72	Biasa	Bahwa ter Ya		Menuntut Supaya M	Mengadili	lv penggelapan	2 tahun
7	618/Pid.S	20-Dec-16	Perlindungan Ana	Daniel Kri pondius wonda alias	Pemberti	99	Biasa	bahwa ter Ya		Menuntut Supaya M	Mengadili	lv perlindungan ana	10 tahun dan 6 bulan
8	605/Pid.S	19-Dec-16	Lain-lain	Jumadi SH agus sumanta alias ati	Minutasi	63	Biasa	Mengadili Ya		Menuntut Supaya M	Mengadili	lv lain-lain	4 bulan
9	606/Pid.S	19-Dec-16	Lain-lain	Jumadi SH nunung tri susantp bi	Minutasi	57	Biasa	Pertama : Ya		Menuntut Supaya M	Mengadili	lv lain-lain	4 bulan
10	607/Pid.B	19-Dec-16	Penadahan , Pene	Hanifah St niko alit prakoso alias	Minutasi	44	Biasa	Bahwa ter Ya		Menuntut Supaya M	Mengadili	lv penadahan, pene	6 bulan
11	608/Pid.B	19-Dec-16	Pencurian	Daru Trias arif nugroho alias gep	Minutasi	37	Biasa	Bahwa ter Ya		Menuntut Menyatak	Mengadili	lv pencurian	7 bulan
12	609/Pid.B	19-Dec-16	Pencurian	Andy Nug josef rianto	Minutasi	43	Biasa	KESATU : E Ya		Menuntut upaya Ma	Mengadili	lv pencurian	3 bulan dan 15 hari
13	610/Pid.B	19-Dec-16	Kejahatan Perjudi	Handry Sumursito alias nggesit	Minutasi	49	Biasa	KESATU : E Ya		Menuntut Supaya M	Mengadili	lv perjudian	6 bulan
14	611/Pid.B	19-Dec-16	Pencurian	Handry Su budi priyo leksono ali	Minutasi	77	Biasa	KESATU : E Ya		Menuntut Supaya M	Mengadili	lv pencurian	10 bulan
15	612/Pid.B	19-Dec-16	Pencurian	Arifah Mi rajul alias firman bin	Minutasi	57	Biasa	Bahwa Te Ya		Menuntut Supaya M	Mengadili	lv pencurian	6 bulan
16	604/Pid.S	14-Dec-16	Narkotika	Kusuma Jreno jurdan faizi bin v	Minutasi	54	Biasa	bahwa ter Ya		Menuntut Supaya M	Mengadili	lv narkotika	5 bulan
17	600/Pid.B	14-Dec-16	Pencurian	Anita Kaja bogi ernanto alias boq	Minutasi	75	Biasa	Kesatu : E Ya		Menuntut Supaya M	Mengadili	lv penggelapan	8 bulan
18	601/Pid.B	14-Dec-16	Pencurian	Asep Purc riyari sugeng riyadi bi	Minutasi	34	Biasa	Bahwa ter Ya		Menuntut Menyatak	Mengadili	lv pencurian	10 bulan
19	602/Pid.B	14-Dec-16	Kejahatan Perjudi	Dhudi Hac agus kalimantono alia	Minutasi	54	Biasa	Kesatu : B Ya		Menuntut Menyatak	Mengadili	lv perjudian	1 tahun
20	603/Pid.S	14-Dec-16	Informasi dan Trar	Heri Supri mardha dwi angara a	Minutasi	64	Biasa	Bahwa Ter Ya		Menuntut Menyatak	Mengadili	lv Informasi dan Trar	2 tahun dan 3 bulan
21	592/Pid.B	13-Dec-16	Pencurian	Dian Susai robby febrion alias ro	Minutasi	42	Biasa	Bahwa ter Ya		Menuntut Supaya M	Mengadili	lv pencurian	6 bulan
22	593/Pid.B	13-Dec-16	Pencurian	Dian Susai imam choiri alias imai	Minutasi	71	Biasa	Bahwa ter Ya		Menuntut Supaya M	Mengadili	lv pencurian	8 bulan

Link berkas asli:

http://tiny.cc/TA_Fauziyah



Lampiran 2 Script dan Output Topic Modeling

```
#membuka file
import pandas as pd
data = pd.read_excel('dataputusan.xlsx')
data.head()
```

	text
0	Mengadili Menyatakan Terdakwa 1. WALIJO alias ...
1	Mengadili Menyatakan Terdakwa 1. WIDIYONO bin ...
2	Mengadili Menyatakan terdakwa RAHMAT HIDAYAT ...
3	Mengadili Menyatakan terdakwa RENDY CAHYO NUGR...
4	Mengadili Menyatakan terdakwa Purwanta als. Pu...

```
#Lower Casing
data['text'] = data['text'].apply(lambda x: " ".join(x.lower() for x in x.split()))
data['text'].head()
```

0	mengadili menyatakan terdakwa 1. walijo alias ...
1	mengadili menyatakan terdakwa 1. widiyono bin ...
2	mengadili menyatakan terdakwa rahmat hidayat a...
3	mengadili menyatakan terdakwa rendy cahyo nugn...
4	mengadili menyatakan terdakwa purwanta als. pu...

Name: text, dtype: object

```
# Removing Punctuation
data['text'] = data['text'].str.replace('[^\w\s]','')
data['text'].head()
```

0	mengadili menyatakan terdakwa 1 walijo alias j...
1	mengadili menyatakan terdakwa 1 widiyono bin k...
2	mengadili menyatakan terdakwa rahmat hidayat a...
3	mengadili menyatakan terdakwa rendy cahyo nugn...
4	mengadili menyatakan terdakwa purwanta als pur...

Name: text, dtype: object

```
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory
```

```
factory = StopWordRemoverFactory()
stopwords = factory.get_stop_words()
print(stopwords)
```

```
['yang', 'untuk', 'pada', 'ke', 'para', 'namun', 'menurut', 'antara', 'dia', 'dua', 'ia', 'seperti', 'jika', 'jika', 'sehing
a', 'kembali', 'dan', 'tidak', 'ini', 'karena', 'kepada', 'oleh', 'saat', 'harus', 'sementara', 'setelah', 'belum', 'kami', 'se
kita', 'bagi', 'serta', 'di', 'dari', 'telah', 'sebagai', 'masih', 'hal', 'ketika', 'adalah', 'itu', 'dalam', 'bisa', 'bahwa',
'atau', 'hanya', 'kita', 'dengan', 'akan', 'juga', 'ada', 'mereka', 'sudah', 'saya', 'terhadap', 'secara', 'agar', 'lain', 'and
a', 'begitu', 'mengapa', 'kenapa', 'yaitu', 'yakni', 'daripada', 'itulah', 'lagi', 'maka', 'tentang', 'demi', 'dimana', 'keman
a', 'pula', 'sambil', 'sebelum', 'sesudah', 'supaya', 'guna', 'kah', 'pun', 'sampai', 'sedangkan', 'selagi', 'sementara', 'teta
pi', 'apakah', 'kecuali', 'sebab', 'selain', 'seolah', 'seraya', 'seterusnya', 'tanpa', 'agak', 'boleh', 'dapat', 'dsb', 'dst',
'dll', 'dahulu', 'dulunya', 'anu', 'demikian', 'tapi', 'ingin', 'juga', 'nggak', 'mari', 'nanti', 'melainkan', 'oh', 'ok', 'seh
arusnya', 'sebetulnya', 'setiap', 'setidaknya', 'sesuatu', 'pasti', 'saja', 'toh', 'ya', 'walau', 'tolong', 'tentu', 'amat', 'a
palagi', 'bagaimanapun']
```

```
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory
```

```
factory = StopWordRemoverFactory()
stopword = factory.get_stop_words()
```

```
# Kalimat
data['text'] = data['text'].apply(lambda x: " ".join(x for x in x.split() if x not in stopword))
data['text'].head()
```

0	mengadili menyatakan terdakwa 1 walijo alias j...
1	mengadili menyatakan terdakwa 1 widiyono bin k...
2	mengadili menyatakan terdakwa rahmat hidayat a...
3	mengadili menyatakan terdakwa rendy cahyo nugn...
4	mengadili menyatakan terdakwa purwanta als pur...

Name: text, dtype: object

```
stopword

['yang',
 'untuk',
 'pada',
 'ke',
 'para',
 'namun',
 'menurut',
 'antara',
 'dia',
 'dua',
 'ia',
 'seperti',
 'jika',
 'jika',
 'sehingga',
 'kembali',
 'dan',
 'tidak',
 'ini',
 '.....']
```

```
import requests
def stopwords():
    r = requests.get("https://raw.githubusercontent.com/masdevid/ID-Stopwords/master/id.stopwords.02.01.2016.txt").text
    data = []
    for x in r.split("\n"):
        data.append(x)
    return data
```

```
stopwords()

['ada',
 'adalah',
 'adanya',
 'adapun',
 'agak',
 'agaknya',
 'agar',
 'akan',
 'akankah',
 'akhir',
 'akhiri',
 'akhirnya',
 'aku',
 'akulah',
 'amat',
 'amatlah',
 'anda',
 'andalah',
 'antar',
 '.....']
```

```
# Import Stopword Factory Class
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory

#Create factory
factory = StopWordRemoverFactory()
more_stopword = ['sekira', 'coba', 'berdasarkan', 'menyimpan', 'menguasai', 'bukan', 'lalu', 'caracara', 'berikut', 'setidak', 'waktu',
 'selanjutnya', 'disebut', 'kemudian', 'bin', 'alm', 'beserta', 'alias', 'buah', 'masingmasing', 'berisi', 'almarhum',
 'binti', 'tanggal', 'no', 'an', 'mengadili', 'menyatakan', 'als', 'hari', 'jam', 'pengadilan', 'negeri', 'yogyakarta',
 'wib', 'kesatu', 'kedua', 'ketiga', 'pertama', 'jam', 'kabupaten', 'bahwa', 'memutuskan', 'tindak', 'pidana',
 'majelis', 'hakim', 'sleman', 'tendakwa', 'alamat', 'terbuat', 'huruf']

# Tambahkan Stopword Baru
stopwordplus = factory.get_stop_words()+stopwords()+more_stopword

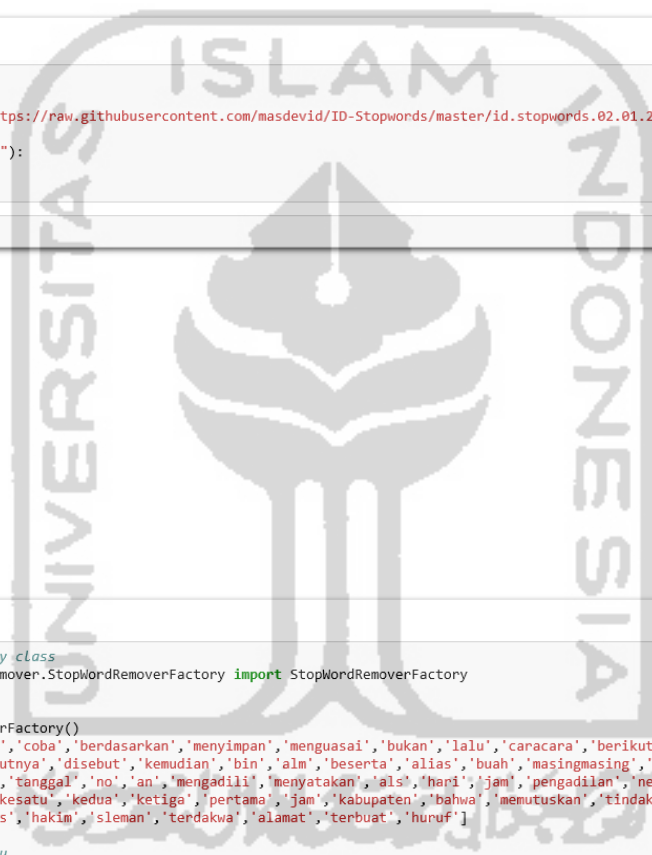
data['text'] = data['text'].apply(lambda x: " ".join(x for x in x.split() if x not in stopwordplus))
data['text'].head()
```

```
0    1 walijo jo 2 ferdyanto terbukti sah bersalah ...
1    1 widiyono kertoarjo 2 saptana harjo suwito te...
2    rahmat hidayat kencut bahrudin terbukti sah be...
3    rendy cahyo nugroho sapatra rendy rakiman terb...
4    purwanta pur mitro wiyono terbukti sah bersala...
Name: text, dtype: object
```

```
text = data['text']
text_list = [i.split() for i in text]
```

```
print(len(text_list))
```

2198



```
print(text_list)

[['1', 'waliyo', 'jo', '2', 'ferdyanto', 'terbukti', 'sah', 'bersalah', 'hak', 'sengaja', 'kesempatan', 'khalayak', 'permaina', 'judi', 'menjatuhkan', 'penjara', '4', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'menetapkan', 'barang', 'bukti', '1', 'handphone', 'merk', 'xiomi', 'warna', 'hitam', 'putih', '1', 'handphone', 'merk', 'nokia', 'warna', 'biru', '1', 'handphone', 'merk', 'sony', '1', 'lembar', 'kertas', 'rekaman', 'pembelian', 'judi', 'togel', 'hk', '1', 'buku', 'catatan', 'rekapan', '1', 'lembar', 'catatan', 'angka', '1', 'lembar', 'catatan', 'angka', '2', 'buku', 'ramalan', 'dirampas', 'dimusnahkan', 'uang', 'tunai', 'rp', '1200000', 'negara', 'membebakan', 'biaya', 'perkara', 'rp', '20000', 'ribu', 'rupiah'], ['1', 'widiyono', 'kertoarjo', '2', 'saptana', 'harjo', 'suwito', 'terbukti', 'sah', 'bersalah', 'hak', 'sengaja', 'kesempatan', 'khalayak', 'permainan', 'judi', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'barang', 'bukti', '1', 'handphone', 'merk', 'cross', 'seri', 'e11t', 'warna', 'hitam', '1', 'sim', 'card', '085743406614', '1', 'hand', 'phone', 'merk', 'nokia', 'seri', '1112', 'warna', 'biru', 'hitam', '1', 'simcard', '081904219304', '2', 'lembar', 'rekapan', 'penjualan', 'judi', 'togel', 'dirampas', 'dimusnahkan', 'uang', 'tunai', 'rp', '9500000', 'uang', 'tunai', 'rp', '100000000', 'dirampas', 'negara', 'membebakan', 'biaya', 'perkara', 'rp', '200000', 'ribu', 'rupiah'], ['rahmat', 'hidayat', 'kencut', 'bahrudin', 'terbukti', 'sah', 'bersalah', 'kekerasan', 'orang', 'menyebabkan', 'orang', 'luka', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'membebakan', 'membayar', 'biaya', 'perkara', 'rp', '200000', 'ribu', 'rupiah'], ['rendy', 'cahyo', 'nugroho', 'saputra', 'rendy', 'rakiman', 'terbukti', 'sah', 'bersalah', 'penggelapan', 'orang', 'barang', 'hubungan', 'kerja', 'berlanjut', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'menetapkan', 'barang', 'bukti', 'uang', 'rp', '72000', 'tujuh', 'puluh', 'ribu', 'rupiah', 'rincian', 'lembar', 'uang', 'pecahan', 'rp50000', 'puluh', 'r
```

```
import re

# Function to Tokenize words
def tokenize(text):
    tokens = re.split('\W+', text) #\W+ means that either a word character (A-Za-z0-9) or a dash (-) can go there.
    return tokens

data['text'] = data['text'].apply(lambda x: tokenize(x.lower()))
#We convert to lower as Python is case-sensitive.
data.head()
```

	text
0	[1, 'waliyo', 'jo', '2', 'ferdyanto', 'terbukti', 'sah', 'bersalah', 'hak', 'sengaja', 'kesempatan', 'khalayak', 'permainan', 'judi', 'menjatuhkan', 'penjara', '4', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'menetapkan', 'barang', 'bukti', '1', 'handphone', 'merk', 'xiomi', 'warna', 'hitam', 'putih', '1', 'handphone', 'merk', 'nokia', 'warna', 'biru', '1', 'handphone', 'merk', 'sony', '1', 'lembar', 'kertas', 'rekaman', 'pembelian', 'judi', 'togel', 'hk', '1', 'buku', 'catatan', 'rekapan', '1', 'lembar', 'catatan', 'angka', '1', 'lembar', 'catatan', 'angka', '2', 'buku', 'ramalan', 'dirampas', 'dimusnahkan', 'uang', 'tunai', 'rp', '1200000', 'negara', 'membebakan', 'biaya', 'perkara', 'rp', '20000', 'ribu', 'rupiah'], ['1', 'widiyono', 'kertoarjo', '2', 'saptana', 'harjo', 'suwito', 'terbukti', 'sah', 'bersalah', 'hak', 'sengaja', 'kesempatan', 'khalayak', 'permainan', 'judi', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'menetapkan', 'barang', 'bukti', 'uang', 'rp', '72000', 'tujuh', 'puluh', 'ribu', 'rupiah', 'rincian', 'lembar', 'uang', 'pecahan', 'rp50000', 'puluh', 'r
1	[1, 'widiyono', 'kertoarjo', '2', 'saptana', 'harjo', 'suwito', 'terbukti', 'sah', 'bersalah', 'hak', 'sengaja', 'kesempatan', 'khalayak', 'permainan', 'judi', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'menetapkan', 'barang', 'bukti', 'uang', 'rp', '72000', 'tujuh', 'puluh', 'ribu', 'rupiah', 'rincian', 'lembar', 'uang', 'pecahan', 'rp50000', 'puluh', 'r
2	['rahmat', 'hidayat', 'kencut', 'bahrudin', 'terbukti', 'sah', 'bersalah', 'kekerasan', 'orang', 'menyebabkan', 'orang', 'luka', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'membebakan', 'membayar', 'biaya', 'perkara', 'rp', '200000', 'ribu', 'rupiah'], ['rendy', 'cahyo', 'nugroho', 'saputra', 'rendy', 'rakiman', 'terbukti', 'sah', 'bersalah', 'penggelapan', 'orang', 'barang', 'hubungan', 'kerja', 'berlanjut', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'menetapkan', 'barang', 'bukti', 'uang', 'rp', '72000', 'tujuh', 'puluh', 'ribu', 'rupiah', 'rincian', 'lembar', 'uang', 'pecahan', 'rp50000', 'puluh', 'r
3	['rendy', 'cahyo', 'nugroho', 'saputra', 'rendy', 'rakiman', 'terbukti', 'sah', 'bersalah', 'penggelapan', 'orang', 'barang', 'hubungan', 'kerja', 'berlanjut', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'menetapkan', 'barang', 'bukti', 'uang', 'rp', '72000', 'tujuh', 'puluh', 'ribu', 'rupiah', 'rincian', 'lembar', 'uang', 'pecahan', 'rp50000', 'puluh', 'r
4	['purwanta', 'pur', 'mitro', 'wiyono', 'terbukti', 'sah', 'bersalah', 'hak', 'sengaja', 'kesempatan', 'khalayak', 'permainan', 'judi', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'menetapkan', 'barang', 'bukti', 'uang', 'rp', '72000', 'tujuh', 'puluh', 'ribu', 'rupiah', 'rincian', 'lembar', 'uang', 'pecahan', 'rp50000', 'puluh', 'r

```
data.to_csv('putusanlimit11.csv', sep=',')
import pandas as pd
data = pd.read_csv('putusanlimit11.csv')
data.head()
```

	Unnamed: 0	text
0	0	[1, 'waliyo', 'jo', '2', 'ferdyanto', 'terbukti', 'sah', 'bersalah', 'hak', 'sengaja', 'kesempatan', 'khalayak', 'permainan', 'judi', 'menjatuhkan', 'penjara', '4', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'menetapkan', 'barang', 'bukti', '1', 'handphone', 'merk', 'xiomi', 'warna', 'hitam', 'putih', '1', 'handphone', 'merk', 'nokia', 'warna', 'biru', '1', 'handphone', 'merk', 'sony', '1', 'lembar', 'kertas', 'rekaman', 'pembelian', 'judi', 'togel', 'hk', '1', 'buku', 'catatan', 'rekapan', '1', 'lembar', 'catatan', 'angka', '1', 'lembar', 'catatan', 'angka', '2', 'buku', 'ramalan', 'dirampas', 'dimusnahkan', 'uang', 'tunai', 'rp', '1200000', 'negara', 'membebakan', 'biaya', 'perkara', 'rp', '20000', 'ribu', 'rupiah'], ['1', 'widiyono', 'kertoarjo', '2', 'saptana', 'harjo', 'suwito', 'terbukti', 'sah', 'bersalah', 'hak', 'sengaja', 'kesempatan', 'khalayak', 'permainan', 'judi', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'menetapkan', 'barang', 'bukti', 'uang', 'rp', '72000', 'tujuh', 'puluh', 'ribu', 'rupiah', 'rincian', 'lembar', 'uang', 'pecahan', 'rp50000', 'puluh', 'r
1	1	[1, 'widiyono', 'kertoarjo', '2', 'saptana', 'harjo', 'suwito', 'terbukti', 'sah', 'bersalah', 'hak', 'sengaja', 'kesempatan', 'khalayak', 'permainan', 'judi', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'menetapkan', 'barang', 'bukti', 'uang', 'rp', '72000', 'tujuh', 'puluh', 'ribu', 'rupiah', 'rincian', 'lembar', 'uang', 'pecahan', 'rp50000', 'puluh', 'r
2	2	['rahmat', 'hidayat', 'kencut', 'bahrudin', 'terbukti', 'sah', 'bersalah', 'kekerasan', 'orang', 'menyebabkan', 'orang', 'luka', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'membebakan', 'membayar', 'biaya', 'perkara', 'rp', '200000', 'ribu', 'rupiah'], ['rendy', 'cahyo', 'nugroho', 'saputra', 'rendy', 'rakiman', 'terbukti', 'sah', 'bersalah', 'penggelapan', 'orang', 'barang', 'hubungan', 'kerja', 'berlanjut', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'menetapkan', 'barang', 'bukti', 'uang', 'rp', '72000', 'tujuh', 'puluh', 'ribu', 'rupiah', 'rincian', 'lembar', 'uang', 'pecahan', 'rp50000', 'puluh', 'r
3	3	['rendy', 'cahyo', 'nugroho', 'saputra', 'rendy', 'rakiman', 'terbukti', 'sah', 'bersalah', 'penggelapan', 'orang', 'barang', 'hubungan', 'kerja', 'berlanjut', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'menetapkan', 'barang', 'bukti', 'uang', 'rp', '72000', 'tujuh', 'puluh', 'ribu', 'rupiah', 'rincian', 'lembar', 'uang', 'pecahan', 'rp50000', 'puluh', 'r
4	4	['purwanta', 'pur', 'mitro', 'wiyono', 'terbukti', 'sah', 'bersalah', 'hak', 'sengaja', 'kesempatan', 'khalayak', 'permainan', 'judi', 'menjatuhkan', 'penjara', '1', 'menetapkan', 'penangkapan', 'penahanan', 'dijalani', 'dikurangkan', 'dijatuhkan', 'menetapkan', 'ditahan', 'menetapkan', 'barang', 'bukti', 'uang', 'rp', '72000', 'tujuh', 'puluh', 'ribu', 'rupiah', 'rincian', 'lembar', 'uang', 'pecahan', 'rp50000', 'puluh', 'r

```
from sklearn.feature_extraction.text import CountVectorizer
count_vectorizer = CountVectorizer(encoding='latin-1', ngram_range=(1, 1), tokenizer=None, analyzer = 'word')
countvec = count_vectorizer.fit_transform(data.text).toarray()
countvec

array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       ...,
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]], dtype=int64)
```

```
countvec2 = pd.DataFrame(countvec)
countvec2
```

	0	1	2	3	4	5	6	7	8	9	...	21860	21861	21862	21863	21864	21865	21866	21867	21868	21869	
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0

```
kata_kata = count_vectorizer.get_feature_names()
countvec3 = pd.DataFrame(countvec, columns=kata_kata)
countvec3
```

	00	000	0000	00000	00000000	000004	000006	000014	0001skdirbprimi2011	00027	...	zulyati	zumrotul	zupiter	zura	zusri	zusuki	zwnerbd50415
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0

```
from sklearn.feature_extraction.text import TfidfTransformer
transformer = TfidfTransformer(norm=None, use_idf=True, smooth_idf=False, sublinear_tf=False)
tfidf = transformer.fit_transform(countvec)
tfidf
```

<2198x21870 sparse matrix of type '<class 'numpy.float64''>
with 147856 stored elements in Compressed Sparse Row format>

```
tfidf1 = tfidf.toarray()
tfidf1
```

```
array([[0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       ...,
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.],
       [0., 0., 0., ..., 0., 0., 0.]])
```

```
tfidf2 = pd.DataFrame(tfidf1)
tfidf2
```

	0	1	2	3	4	5	6	7	8	9	...	21860	21861	21862	21863	21864	21865	21866	21867	21868	21869
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

```
kata_kata2 = count_vectorizer.get_feature_names()
df1 = pd.DataFrame(tfidf1, columns=kata_kata2)
df1
```

	00	000	0000	00000	00000000	000004	000006	000014	0001skdirbprrimi2011	00027	...	zulyati	zumrotul	zupiter	zura	zusri	zusuki	zwnerbd5041
0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0
10	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0

```
import numpy as np
tfidf=np.transpose(df1)
```

```
from xlswriter.utility import xl_rowcol_to_cell
saveresult = pd.ExcelWriter('tfidfsoftware.xlsx', engine='xlswriter')
tfidf.to_excel(saveresult, index=False, sheet_name='report')
saveresult.save()
```

```
import numpy as np
konvek=np.transpose(countvec3)
```

```
from xlswriter.utility import xl_rowcol_to_cell
saveresult = pd.ExcelWriter('perhitunganputusanlimit11.xlsx', engine='xlswriter')
konvek.to_excel(saveresult, index=False, sheet_name='report')
saveresult.save()
```

```
#Create Bigram models
from gensim.models import Phrases
# Add bigrams and trigrams to docs, minimum count 10 means only that appear 10 times or more.
unigram = Phrases(text_list, min_count=10)

for idx in range(len(text_list)):
    for token in unigram[text_list[idx]]:
        if '_' in token:
            # Token is a bigram, add to document.
            text_list[idx].append(token)
```

```
from gensim import corpora, models
# Create a dictionary representation of the documents.
dictionary = corpora.Dictionary(text_list)

dictionary.filter_extremes(no_below=5, no_above=0.2)
#no_below (int, optional) - Keep tokens which are contained in at Least no_below documents.
#no_above (float, optional) - Keep tokens which are contained in no more than no_above documents (fraction of total corpus size,
print(dictionary)
```

< Dictionary(3078 unique tokens: ['200000', '200000_ribu', '4', 'angka', 'bersalah_hak']...)


```
#https://radimrehurek.com/gensim/tut1.html
#build corpus

doc_term_matrix = [dictionary.doc2bow(doc) for doc in text_list]
#The function doc2bow converts document (a list of words) into the bag-of-words format

'''The function doc2bow() simply counts the number of occurrences of each distinct word,
converts the word to its integer word id and returns the result as a sparse vector.
The sparse vector [(0, 1), (1, 1)] therefore reads: in the document "Human computer interaction",
the words computer (id 0) and human (id 1) appear once;
the other ten dictionary words appear (implicitly) zero times.'''

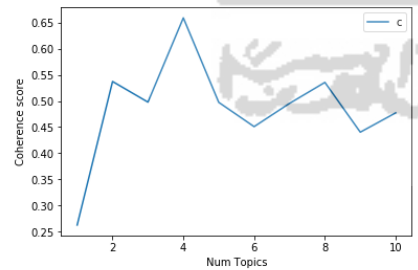
print(len(doc_term_matrix))
print(doc_term_matrix[100])

tfidf = models.TfidfModel(doc_term_matrix) #build TF-IDF model
corpus_tfidf = tfidf[doc_term_matrix]
```

```
2198
[(41, 1), (50, 1), (51, 1), (52, 1), (116, 1), (182, 1), (214, 1), (232, 1), (286, 1), (308, 2), (315, 1), (342, 1), (343, 1),
(385, 1), (394, 1), (395, 1), (398, 1), (399, 1), (406, 1), (413, 1), (444, 1), (445, 1), (449, 1), (483, 1), (515, 1), (540,
1), (583, 2), (585, 2), (586, 1), (592, 1), (600, 1), (601, 1), (609, 2), (817, 3), (821, 1), (824, 1), (826, 3), (834, 1), (83
6, 1), (840, 1), (844, 1), (854, 2), (855, 1), (856, 1), (863, 3), (864, 1), (872, 5), (873, 1), (1039, 1), (1086, 1), (1090,
1), (1098, 1), (1103, 1), (1111, 3), (1141, 2), (1142, 2), (1143, 3), (1144, 2), (1145, 2), (1146, 2), (1147, 2), (1148, 3), (1
151, 3), (1152, 2), (1153, 2), (1160, 1), (1374, 1), (1422, 1), (1426, 1), (1427, 1), (1428, 1), (1429, 2), (1430, 1), (1431,
1), (1432, 1), (1433, 1)]
```

```
from gensim.models.coherencemodel import CoherenceModel
from gensim.models.ldamodel import LdaModel
from gensim.corpora.dictionary import Dictionary
from numpy import array
#function to compute coherence values
def compute_coherence_values(dictionary, corpus, texts, limit, start, step):
    coherence_values = []
    model_list = []
    for num_topics in range(start, limit, step):
        model = LdaModel(corpus=corpus, id2word=dictionary, num_topics=num_topics, iterations=100)
        model_list.append(model)
        coherencemodel = CoherenceModel(model=model, texts=texts, dictionary=dictionary, coherence='c_v')
        coherence_values.append(coherencemodel.get_coherence())
    return model_list, coherence_values
```

```
start=1
limit=11
step=1
model_list, coherence_values = compute_coherence_values(dictionary, corpus=corpus_tfidf,
                                                         texts=text_list, start=start, limit=limit, step=step)
#show graphs
import matplotlib.pyplot as plt
x = range(start, limit, step)
plt.plot(x, coherence_values)
plt.xlabel("Num Topics")
plt.ylabel("Coherence score")
plt.legend(("coherence_values"), loc='best')
plt.show()
```



```
# Print the coherence scores
for m, cv in zip(x, coherence_values):
    print("Num Topics =", m, " has Coherence Value of", round(cv, 5))
```

```
Num Topics = 1 has Coherence Value of 0.26257
Num Topics = 2 has Coherence Value of 0.53723
Num Topics = 3 has Coherence Value of 0.49773
Num Topics = 4 has Coherence Value of 0.65895
Num Topics = 5 has Coherence Value of 0.49732
Num Topics = 6 has Coherence Value of 0.45086
Num Topics = 7 has Coherence Value of 0.49529
Num Topics = 8 has Coherence Value of 0.53551
Num Topics = 9 has Coherence Value of 0.44025
Num Topics = 10 has Coherence Value of 0.47719
```

```
from pprint import pprint

model = LdaModel(corpus=corpus_tfidf, id2word=dictionary, random_state=0, num_topics=4)
pprint(model.print_topics())
```

```
[(0,
 '0.003*memberatkan" + 0.003*stnk" + 0.003*ab" + 0.003*vario" + '
 '0.003*rp_2000" + 0.003*memberatkan_menjatuhkan" + 0.003*2016" + '
 '0.003*ii" + 0.003*honda" + 0.003*surat'),
 (1,
 '0.009*disamakan" + 0.004*shabu" + 0.004*kaca" + 0.003*pipet" + '
 '0.003*atm" + 0.003*pipet_kaca" + 0.003*kartu" + 0.003*plastik" + '
 '0.003*iman" + 0.003*kartu_atm'),
 (2,
 '0.006*kayu" + 0.005*rehabilitasi" + 0.004*agus" + 0.004*senjata" + '
 '0.004*cm" + 0.004*pedang" + 0.003*ade" + 0.003*bilah" + 0.003*rp_2000" + '
 '+ 0.003*besi'),
 (3,
 '0.004*2016" + 0.004*uang" + 0.004*simcard" + 0.004*tunai" + '
 '0.004*uang_tunai" + 0.004*judi" + 0.004*togel" + 0.004*nomor_simcard" + '
 '0.004*hp" + 0.004*kertas')]
```

```
model = LdaModel(corpus=corpus_tfidf, id2word=dictionary, random_state=0, num_topics=4)
```

```
for idx, topic in model.print_topics(-1):
    print('Topic: {} Word: {}'.format(idx, topic))
```

```
Topic: 0 Word: 0.003*memberatkan" + 0.003*stnk" + 0.003*ab" + 0.003*vario" + 0.003*rp_2000" + 0.003*memberatkan_menjatuhkan" + 0.003*2016" + 0.003*ii" + 0.003*honda" + 0.003*surat"
Topic: 1 Word: 0.009*disamakan" + 0.004*shabu" + 0.004*kaca" + 0.003*pipet" + 0.003*atm" + 0.003*pipet_kaca" + 0.003*kartu" + 0.003*plastik" + 0.003*iman" + 0.003*kartu_atm"
Topic: 2 Word: 0.006*kayu" + 0.005*rehabilitasi" + 0.004*agus" + 0.004*senjata" + 0.004*cm" + 0.004*pedang" + 0.003*ade" + 0.003*bilah" + 0.003*rp_2000" + 0.003*besi"
Topic: 3 Word: 0.004*2016" + 0.004*uang" + 0.004*simcard" + 0.004*tunai" + 0.004*uang_tunai" + 0.004*judi" + 0.004*togel" + 0.004*nomor_simcard" + 0.004*hp" + 0.004*kertas"
```

```
import pandas as pd
top_words_per_topic = []
for t in range(model.num_topics):
    top_words_per_topic.extend([(t, ) + x for x in model.show_topic(t, topn = 10)])

#pd.DataFrame(top_words_per_topic, columns=['Topic', 'Word', 'P']).to_csv("top_words.csv")
df = pd.DataFrame(top_words_per_topic, columns=['Topic', 'Word', 'P']).to_csv("top_words_limit11.csv")
print(df)
```

None

```
import gensim
import pyLDAvis.gensim;pyLDAvis.enable_notebook()

data = pyLDAvis.gensim.prepare(model, corpus_tfidf, dictionary)
print(data)
pyLDAvis.save_html(data, 'lda-gensimputusan11.html')
```

PreparedData	topic	topic_coordinates=	Freq	cluster	topics	x	y	topic_info=	Category	Freq	Term	Tota
3	32.225689	1	1	-0.083662	-0.002306							
0	31.699133	1	2	0.040118	-0.064757							
1	21.369488	1	3	0.002150	0.009455							
2	14.705688	1	4	0.041394	0.057608							
1	loglift	logprob										
term												
1272	Default	22.000000		disamakan	22.000000	30.0000	30.0000					
1025	Default	16.000000		kayu	16.000000	29.0000	29.0000					
604	Default	11.000000		rehabilitasi	11.000000	28.0000	28.0000					
35	Default	16.000000		togel	16.000000	27.0000	27.0000					
1218	Default	11.000000		senjata	11.000000	26.0000	26.0000					
1215	Default	9.000000		pedang	9.000000	25.0000	25.0000					
15	Default	16.000000		judi	16.000000	24.0000	24.0000					
2420	Default	7.000000		ade	7.000000	23.0000	23.0000					
685	Default	14.000000		hongkong	14.000000	22.0000	22.0000					
708	Default	8.000000		iman	8.000000	21.0000	21.0000					
1637	Default	9.000000		bilah	9.000000	20.0000	20.0000					
4730	Default	7.000000			7.000000	19.0000	19.0000					

Lampiran 3 Topic Coherence limit 21, 31, 41, dan 51

Tabel topic coherence limit 21

<i>Num Topics</i>	<i>Coherence Value</i>	<i>Num Topics</i>	<i>Coherence Value</i>
1	0,26257	11	0,52618
2	0,627	12	0,4872
3	0,47109	13	0,44635
4	0,39855	14	0,5282
5	0,39731	15	0,50229
6	0,53608	16	0,53219
7	0,51336	17	0,52723
8	0,52062	18	0,52518
9	0,43365	19	0,50331
10	0,47533	20	0,55406

Tabel topic coherence limit 31

<i>Num Topics</i>	<i>Coherence Value</i>	<i>Num Topics</i>	<i>Coherence Value</i>
1	0,26257	16	0,54436
2	0,52564	17	0,49046
3	0,38249	18	0,51797
4	0,47983	19	0,48745
5	0,55883	20	0,54813
6	0,47815	21	0,49996
7	0,51891	22	0,51992
8	0,56752	23	0,50356
9	0,50139	24	0,5166
10	0,45105	25	0,53346
11	0,46056	26	0,52144
12	0,51988	27	0,50234
13	0,5006	28	0,48034

<i>Num Topics</i>	<i>Coherence Value</i>	<i>Num Topics</i>	<i>Coherence Value</i>
14	0,52063	29	0,51399
15	0,50019	30	0,49858

Tabel topic coherence limit 41

<i>Num Topics</i>	<i>Coherence Value</i>	<i>Num Topics</i>	<i>Coherence Value</i>
1	0,26257	21	0,55757
2	0,57113	22	0,5203
3	0,57219	23	0,51542
4	0,38421	24	0,49631
5	0,43054	25	0,49811
6	0,50027	26	0,49114
7	0,45643	27	0,48895
8	0,52772	28	0,53644
9	0,47434	29	0,51916
10	0,49242	30	0,5007
11	0,45606	31	0,49457
12	0,51052	32	0,50929
13	0,46683	33	0,50161
14	0,52552	34	0,48752
15	0,5324	35	0,50815
16	0,496	36	0,49847
17	0,52632	37	0,51196
18	0,52506	38	0,5076
19	0,50408	39	0,50364
20	0,52078	40	0,49281