

**IMPLEMENTASI *LATENT DIRICHLET ALLOCATION*
(LDA) UNTUK PEMODELAN TOPIK FAKTOR
PERCERAIAN**

(Studi Kasus : Dokumen Putusan Perceraian di Pengadilan Agama Yogyakarta
Tahun 2020)

TUGAS AKHIR

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



Disusun Oleh:
Elen Prihana Putri
18611116

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2022**

**HALAMAN PERSETUJUAN PEMBIMBING
TUGAS AKHIR**

Judul : Implementasi *Latent Dirichlet Allocation* (LDA)
untuk Pemodelan Topik Faktor Perceraian
(Studi Kasus: Dokumen Putusan Perceraian di
Pengadilan Agama Yogyakarta Tahun 2020)

Nama Mahasiswa : Elen Prihana Putri

NIM : 18611116

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

Yogyakarta, 11 April 2022

Mengetahui,

Pembimbing

Ketua Prodi Statistika

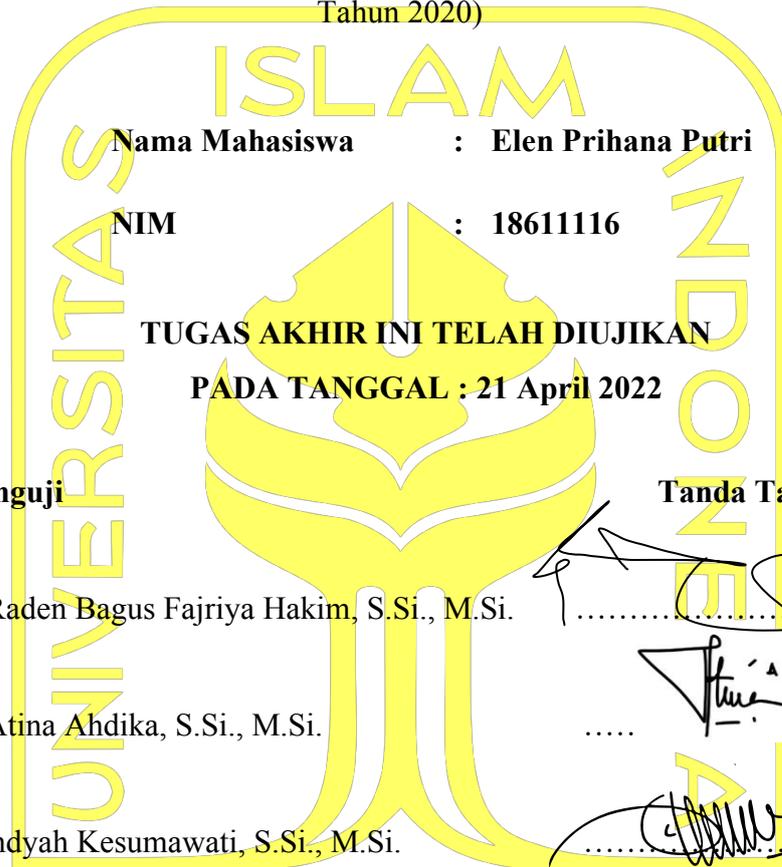

(Ayundyah Kesumawati, S.Si., M.Si.)


(Dr. Edy Widodo, S.Si., M.Si.)

HALAMAN PENGESAHAN
TUGAS AKHIR

**IMPLEMENTASI *LATENT DIRICHLET ALLOCATION* (LDA) UNTUK
PEMODELAN TOPIK FAKTOR PERCERAIAN**

(Studi Kasus : Dokumen Putusan Perceraian di Pengadilan Agama Yogyakarta
Tahun 2020)



Nama Mahasiswa : Elen Prihana Putri

NIM : 18611116

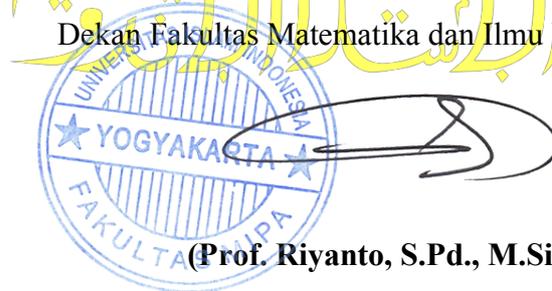
**TUGAS AKHIR INI TELAH DIUJIKAN
PADA TANGGAL : 21 April 2022**

Nama Penguji

Tanda Tangan

1. Dr. Raden Bagus Fajriya Hakim, S.Si., M.Si. 
2. Dr. Atina Ahdika, S.Si., M.Si. 
3. Ayundyah Kesumawati, S.Si., M.Si. 

Mengetahui,
Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



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

KATA PENGANTAR



Assalamu'alaikum Wr.Wb

Alhamdulillah, puji dan syukur penulis panjatkan rasa syukur bagi Allah SWT yang mana berkat limpahan rahmat, hidayah dan karunia-Nya yang telah memberikan keimanan, kekuatan, kesabaran, kelancaran serta keselamatan. Shalawat serta salam penulis haturkan kepada Nabi Muhammad SAW beserta keluarga dan para pengikut-pengikutnya. Oleh karena itu, pada kesempatan ini penulis telah menyelesaikan laporan Tugas Akhir dengan membawakan judul “**Implementasi *Latent Dirichlet Allocation (LDA)* untuk Pemodelan Topik Faktor Perceraian** (Studi Kasus : Dokumen Putusan Perceraian di Pengadilan Agama Yogyakarta Tahun 2020)”.

Dalam penulisan laporan Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana di Program Studi Statistika, Universitas Islam Indonesia. Selama penyusunan laporan, penulis telah banyak mendapat bimbingan dan bantuan dari berbagai pihak. Pada kesempatan ini, penulis mengutarakan rasa syukur dan terima kasih yang sedalam-dalamnya dan tidak terhingga kepada:

1. Bapak Prof. Riyanto, S.Pd., M.Si., Ph.D., selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Indonesia, Yogyakarta.
2. Bapak Dr. Edy Widodo, S.Si., M.Si., selaku Ketua Jurusan Program Studi Statistika beserta jajarannya.
3. Ibu Ayundyah Kesumawati, S.Si., M.Si., selaku Dosen Pembimbing Tugas Akhir, yang telah memberikan waktu luang, ilmu, selalu sabar, dan terus memberi semangat serta motivasi dalam membimbing penulis selama penyusunan Tugas Akhir.
4. Ibu Dra. Ulil Uswah, M.H., selaku pendamping penelitian di Pengadilan Agama Yogyakarta yang telah membantu dan memperlancar dalam persiapan data laporan Tugas Akhir.
5. Bapak serta Ibu Dosen Program Studi Statistika, Universitas Islam Indonesia yang telah memberikan banyak bekal ilmu dan membagikan banyak cerita serta pengalaman selama perkuliahan kepada penulis.

6. Kedua orang tua tercinta yang selalu memberikan doa, motivasi, kasih sayang serta dukungan moral dan materi kepada penulis selama proses perkuliahan dari awal hingga akhir. Tak lupa adik tersayang yang selalu memberikan semangat kepada penulis untuk menjadi pribadi yang lebih baik.
7. Teman-teman terbaik yang telah membantu, mendorong dan mengingatkan penulis untuk selalu semangat dan optimis dalam menyelesaikan penulisan.
8. Seluruh teman-teman Statistika angkatan 2018 yang telah memberikan banyak pengalaman dan informasi dalam perkuliahan serta berjuang bersama dalam masa perkuliahan.
9. Kepada semua pihak yang telah berkenan memberikan bantuan dan dorongan serta kerja sama yang baik, sehingga Tugas Akhir ini dapat diselesaikan dengan lancar.

Penulis menyadari bahwa laporan ini masih jauh dari sempurna, maka dari itu Penulis berharap bisa menerima dan mengaplikasikan berbagai kritik dan saran yang kiranya dapat membangun laporan ini. Semoga penyusunan laporan ini dapat bermanfaat bagi semua pihak yang membutuhkan. Semoga penyusunan Tugas Akhir ini dapat bermanfaat bagi penulis dan bagi semua yang membutuhkan laporan ini nantinya.

Yogyakarta, 11 April 2022



Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN PEMBIMBING TUGAS AKHIR.....	ii
HALAMAN PENGESAHAN TUGAS AKHIR	iii
KATA PENGANTAR.....	iv
DAFTAR ISI	vi
DAFTAR TABEL	viii
DAFTAR GAMBAR.....	ix
DAFTAR LAMPIRAN	x
PERNYATAAN	xi
INTISARI	xii
ABSTRACT	xiii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah	4
1.3. Batasan Masalah.....	4
1.4. Jenis Penelitian dan Metode Analisis.....	5
1.5. Tujuan Penelitian	5
BAB II TINJAUAN PUSTAKA	7
BAB III LANDASAN TEORI	11
3.1. Perceraian.....	11
3.2. Alasan Perceraian.....	11
3.3. Macam-Macam Perceraian.....	12
3.4. Statistika Deskriptif.....	14
3.5. <i>Artificial Intelligence</i> (AI).....	15
3.6. <i>Text Mining</i>	16
3.7. <i>Text PreProcessing</i>	16
3.8. Pembobotan Kata (<i>Term Weighting</i>).....	18
3.8.1 <i>Term Frequency</i> (TF)	18
3.8.2 <i>Inverse Document Frequency</i> (IDF).....	19
3.8.3 <i>Term Frequency-Inverse Document Frequency</i> (TF-IDF).....	19
3.9. <i>Topic Modeling</i>	20
3.10. <i>Latent Dirichlet Allocation</i> (LDA).....	21
3.11. <i>Topic Coherence</i>	23
3.12. <i>Word Cloud</i>	25
BAB IV METODOLOGI PENELITIAN.....	26
4.1. Populasi Penelitian	26
4.2. Tempat dan Waktu Penelitian	26
4.3. Variabel Penelitian	26
4.4. Teknik Sampling	27
4.5. Alat dan Cara Organisir Data.....	27
BAB V HASIL DAN PEMBAHASAN	29
5.1. Analisis Deskriptif	29
5.2. <i>Text Preprocessing</i>	30
5.2.1 <i>Lowercasing</i>	32
5.2.2 <i>Remove Punctuation</i>	34

5.2.3	<i>Stopword Removal</i>	36
5.2.4	<i>Lemmatization</i>	38
5.2.5	<i>Tokenizing</i>	39
5.3.	Pembobotan kata	41
5.4.	<i>Topic Modeling</i> dengan <i>Latent Dirichlet Allocation (LDA)</i>	43
5.4.1	Model LDA Topik ke-1.....	47
5.4.2	Model <i>LDA</i> Topik ke-2	48
5.4.3	Model <i>LDA</i> Topik ke-3	49
BAB VI PENUTUP		52
5.5.	Kesimpulan	52
5.6.	Saran.....	52
DAFTAR PUSTAKA.....		54
LAMPIRAN		59

DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Sebelumnya	9
Tabel 4.1 Variabel Penelitian	26
Tabel 5.1 Teks Awal	31
Tabel 5.2 Hasil Proses <i>Lowercasing</i>	32
Tabel 5.3 Hasil Proses <i>Remove Punctuation</i>	34
Tabel 5.4 Hasil Proses <i>Stopword Removal</i>	36
Tabel 5.5 Hasil Proses <i>Lemmatization</i>	38
Tabel 5.6 Hasil Proses <i>Tokenizing</i>	39
Tabel 5.7 Hasil Perhitungan Nilai TF	41
Tabel 5.8 Hasil Perhitungan Nilai TF-IDF.....	42
Tabel 5.9 Hasil Pembobotan Kata Cerai	43
Tabel 5.10 Hasil Nilai <i>Coherence</i>	44
Tabel 5.11 Hasil Nilai PC	45
Tabel 5.12 Model Topik 1	48
Tabel 5.13 Model Topik 2.....	49
Tabel 5.14 Model Topik 3.....	50

DAFTAR GAMBAR

Gambar 3.1 Ilustrasi <i>Topic Modeling</i>	21
Gambar 3.2 Ilustrasi Metode LDA.....	22
Gambar 3.3 Ilustrasi Menghitung <i>Coherence Score</i>	24
Gambar 4.1 Diagram Alir Penelitian.....	28
Gambar 5.1 Persentase Jenis Kasus Perceraian.....	29
Gambar 5.2 Diagram Faktor Penyebab Perceraian	30
Gambar 5.3 <i>Word Cloud</i>	40
Gambar 5.4 Grafik Nilai <i>Coherence</i>	44
Gambar 5.5 Visualisasi Ketiga Topik	46
Gambar 5.6 <i>Word Cloud</i> Ketiga Topik	46
Gambar 5.7 Visualisasi topik ke-1	47
Gambar 5.8 Visualisasi topik ke-2	48
Gambar 5.9 Visualisasi topik ke-3	50

DAFTAR LAMPIRAN

Lampiran 1	59
Lampiran 2	59
Lampiran 3	59
Lampiran 4	59

PERNYATAAN

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

Yogyakarta, 02 Juni 2022



Elen Pihana Putri

INTISARI

IMPLEMENTASI *LATENT DIRICHLET ALLOCATION* (LDA) UNTUK PEMODELAN TOPIK FAKTOR PERCERAIAN

(Studi Kasus : Dokumen Putusan Perceraian di Pengadilan Agama
Yogyakarta Tahun 2020)

Elen Prihana Putri

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

Perkawinan merupakan salah satu ibadah yang harus dipelihara dengan baik agar tercapai sebuah ikatan yang abadi antara seorang pria dengan wanita sebagai suami istri dengan tujuan membentuk keluarga *sakinah, mawaddah, warahmah*. Namun, sering kali tujuan-tujuan dari perkawinan tersebut tidak dapat terwujud dengan baik dan kandas di perjalanan, maka perselisihan dan pertengkaran sering kali terjadi hingga terciptanya perceraian. Fenomena terjadinya perceraian tidak lepas dari berbagai macam faktor penyebab putusnya perkawinan, sehingga menjadi alasan bagi suami ataupun istri untuk mengajukan perceraian. Adapun faktor penyebab perceraian terdapat dalam dokumen putusan perceraian. Tujuan dalam penelitian ini adalah mengimplementasikan *Topic Modeling* dengan metode *Latent Dirichlet Allocation* (LDA) berdasarkan alasan pengajuan perceraian pada dokumen putusan perceraian sehingga didapatkan beberapa topik terkait faktor penyebab perceraian. Penelitian ini difokuskan pada dokumen putusan perceraian di Pengadilan Agama Yogyakarta pada tahun 2020. Dari hasil analisis deskriptif sebagian besar merupakan perkara cerai gugat dengan persentase 78% dibandingkan dengan cerai talak, kemudian faktor penyebab perceraian yang paling mendominasi adalah perselisihan dan pertengkaran terus-menerus disusul dengan meninggalkan pihak lain, ekonomi, murtad, madat, KDRT, dan cacat badan. Hasil analisis pemodelan topik dengan menggunakan metode LDA diperoleh jumlah topik sebanyak 3 dengan nilai *coherence* 0.29627. Sehingga diperoleh topik yaitu topik 1 cenderung membahas mengenai meninggalkan salah satu pihak, topik 2 cenderung membahas masalah ekonomi, dan topik 3 cenderung membahas mengenai perlakuan salah satu pihak.

Kata Kunci : Perceraian, *Topic Modeling*, *Latent Dirichlet Allocation* (LDA).

ABSTRACT

IMPLEMENTATION OF LATENT DIRICHLET ALLOCATION (LDA) FOR MODELING OF DIVORCE FACTORS

(Case Study : Divorce Decision Documents at the Yogyakarta Religious Court in 2020)

Elen Prihana Putri

*Department of Statistics, Faculty of Mathematics and Natural Sciences
Islamic University of Indonesia*

Marriage is one of the worships that must be maintained properly to achieve a lasting bond between a man and a woman as husband and wife to form a sakinah, mawaddah, warahmah families. However, often the goals of the marriage cannot be realized properly and run aground on the way, so disputes and quarrels often occur until the divorce is created. The phenomenon of divorce cannot be separated from various factors that cause marriage break-ups so that it becomes a reason for husbands or wives to file for divorce. The factors that cause divorce are contained in the divorce decision document. The purpose of this study is to implement topic modeling using the Latent Dirichlet Allocation (LDA) method on the judge's decision document so that several topics related to the factors that cause divorce are obtained. This research focuses on divorce decision documents at the Yogyakarta Religious Court in 2020. From the results of the descriptive analysis, most of the cases are divorced cases with a percentage of 78% compared to talak divorces, then the most dominating factors causing divorce are disputes and constant quarrels followed by leaving the other party, economy, apostasy, madat, domestic violence, and disability. The results of the analysis of topic modeling using the LDA method obtained a total of 3 topics with a coherence value of 0.29627. So that the topics obtained are that topic 1 tends to discuss leaving one party, topic 2 tends to discuss economic problems, and topic 3 tends to discuss the treatment of one party.

Keywords: *Divorce, Topic Modeling, Latent Dirichlet Allocation (LDA).*

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Setiap manusia diciptakan sebagai makhluk sosial yang tidak terlepas dari hakikat kehidupan bermasyarakat. Manusia senantiasa saling berhubungan satu sama lain dalam berbagai hal maupun aspek kehidupan. Pada dasarnya setiap manusia mempunyai kelebihan dan kekurangan pada diri masing-masing dan berbeda antara satu dengan yang lain. Untuk melengkapi kekurangan tersebut, dapat diperoleh dari manusia lain yang mempunyai kelebihan yang diinginkan. Salah satu hal yang dinantikan dalam kehidupan adalah menjalin perkawinan sebagai ikatan lahir batin diantara kedua manusia yaitu pria dan wanita untuk melengkapi satu sama lain.

Menurut Pasal 1 Undang-Undang Nomor 1 Tahun 1974 tentang perkawinan, didalamnya telah disebutkan bahwa perkawinan adalah ikatan lahir batin antara seorang pria dan seorang wanita sebagai suami istri dengan tujuan membentuk keluarga atau rumah tangga yang bahagia dan kekal berdasarkan Ketuhanan Yang Maha Esa. Tujuan utama yang sangat diharapkan dari sebuah perkawinan adalah tercapainya kebahagiaan.

Prinsip suatu perkawinan salah satunya adalah menguatkan ikatan antara suami dan istri sehingga perkawinan dapat berlangsung abadi. Oleh karena itu, segala usaha harus dilakukan untuk mempertahankan perkawinan sampai kematian yang memisahkan. Namun, sering kali tujuan-tujuan dari perkawinan tersebut tidak dapat terwujud dengan baik dan kandas di perjalanan. Banyak alasan dalam masalah rumah tangga berujung perselisihan dan pertengkaran yang tidak dapat terselesaikan. Sehingga, membuat ikatan perkawinan tersebut tidak dapat dipertahankan dan tidak ada harapan untuk rukun kembali seperti sebelumnya maka jalan satu-satunya yang dapat dilakukan adalah memutus ikatan perkawinan dengan perceraian.

Perceraian merupakan sebuah kejadian putusnya hubungan suami dan istri dalam suatu rumah tangga akibat perbedaan prinsip yang tidak dapat dipersatukan kembali dengan berbagai cara. Masalah perceraian diatur dengan tegas pada Pasal

39 ayat (1) dan (2) Undang-Undang Nomor 1 Tahun 1974 tentang perkawinan menyatakan: 1). Perceraian hanya dapat dilakukan di depan Sidang Pengadilan setelah Pengadilan yang bersangkutan berusaha dan tidak berhasil mendamaikan kedua belah pihak. 2). Untuk melakukan perceraian harus ada cukup alasan bahwa antara suami istri itu tidak akan dapat rukun sebagai suami istri (Hermawan, 2018).

Pengajuan gugatan perceraian dan permohonan talak bagi yang beragama Islam diajukan di Pengadilan Agama, sedangkan bagi yang beragama non-Islam diajukan di Pengadilan Negeri. Pasal 54 Undang-Undang Nomor 50 Tahun 2009 tentang Perubahan Kedua atas Undang-undang Nomor 7 Tahun 1989 tentang Peradilan Agama dijelaskan bahwa hukum acara yang berlaku pada pengadilan dalam lingkungan peradilan agama adalah hukum acara perdata yang berlaku pada lingkungan peradilan umum, kecuali yang telah diatur secara khusus dalam undang-undang. Pengadilan agama melakukan pemeriksaan dan memutus perceraian apabila terdapat pihak yang mengajukan gugatan perceraian dan permohonan talak. Pada umumnya terdapat dua jenis perceraian yaitu cerai gugat dan cerai talak. Cerai gugat diajukan atas kehendak istri sedangkan cerai talak diajukan atas kehendak suami. Pengadilan Agama Yogyakarta merupakan salah satu lingkungan peradilan yang menangani perkara perceraian, termasuk perkara cerai gugat dan cerai talak.

Masyarakat D.I. Yogyakarta digolongkan sebagai masyarakat paling bahagia di Indonesia. Dalam perhitungan indeks kebahagiaan yang dilakukan oleh Badan Pusat Statistik tahun 2017, Provinsi D.I. Yogyakarta menempati peringkat ke 8 dari 34 provinsi di Indonesia. Hal ini tentunya berdampak besar dalam kehidupan bermasyarakat. Dimensi yang berkontribusi terbesar dalam penyusunan indeks kebahagiaan adalah dimensi kepuasan hidup dalam lingkup sosial, yang membuktikan bahwa manusia senantiasa lebih bahagia ketika saling berhubungan satu sama lain. Salah satu indikator dalam dimensi kepuasan hidup adalah keharmonisan keluarga (BPS, 2017), dari indikator tersebut tentunya berkaitan ke dalam jumlah kasus perceraian yang tergolong rendah.

Kota Yogyakarta merupakan kota yang memiliki kasus perceraian yang tergolong rendah dibandingkan kabupaten lainnya di Yogyakarta. Hal ini

didukung karena proses perceraian yang dianggap lebih rumit oleh masyarakat. Sehingga dapat menekan angka kasus perceraian yang terjadi. Menurut juru bicara Pengadilan Agama Yogyakarta, kabupaten lain berkemungkinan memiliki kasus perceraian yang tinggi dikarenakan proses yang mudah. Sedangkan di Pengadilan Agama Yogyakarta dianggap lebih rumit karena terdapat mediasi, gugatan balik, dan permintaan yang bermacam-macam dari suami dan istri yang mengajukan perkara perceraian (Assidiq, 2019). Masyarakat Kota Yogyakarta merupakan masyarakat yang heterogen dari beraneka ragam budaya dan profesi. Kota Yogyakarta juga termasuk wilayah dengan penduduknya yang memiliki pola pikir modern, sehingga menjadikan penduduknya memiliki kesadaran hukum tentang perceraian. Selain itu, pada tahun 2020 di Kota Yogyakarta bahkan seluruh Indonesia dihadapkan dengan situasi pandemi Covid-19. Masyarakat diharuskan mematuhi protokol kesehatan sesuai aturan yang dibuat oleh pemerintah, salah satunya yaitu dengan pembatasan aktivitas masyarakat.

Melihat kondisi yang terjadi seperti di atas, tentunya membuat masyarakat berpikir kembali sebelum mengajukan perceraian dan seharusnya memiliki alasan yang sangat penting sehingga harus mengajukan perceraian. Oleh sebab itu, penulis tertarik untuk meneliti dan mengkaji fenomena perceraian terutama dalam hal alasan pengadu dalam mengajukan permohonan cerai di Pengadilan Agama Yogyakarta. Adapun alasan pengajuan perceraian terdapat dalam dokumen putusan perceraian bagian duduk perkara, yang sebagian besar mengandung kalimat panjang dan sulit dimengerti oleh masyarakat umum atau pembaca. Berdasarkan permasalahan tersebut peneliti ingin membantu masyarakat umum atau pembaca dalam memahami bahasan yang berasal dari dokumen putusan perceraian tersebut dengan menggunakan analisis *Topic Modeling*.

Perkembangan teknologi yang berkaitan dengan teks dapat dimanfaatkan untuk mendapatkan topik dari kumpulan kalimat menjadikan fokus utama dalam penelitian ini. Salah satu metode yang dapat digunakan untuk mendapatkan beberapa topik yaitu analisis *Topic Modeling*. Peneliti akan melakukan *Topic Modeling* terhadap alasan pengajuan perceraian pada dokumen putusan perceraian di Pengadilan Agama Yogyakarta sehingga didapatkan beberapa topik mengenai faktor penyebab perceraian. *Topic modeling* merupakan sebuah topik yang terdiri

dari kata-kata tertentu yang menyusun topik tersebut, dan dalam satu dokumen memiliki probabilitas masing-masing dari beberapa topik yang dihasilkan (Putra & Kusumawardani, 2017). Beberapa topik yang diperoleh akan divisualisasikan dan dianalisis untuk mendapatkan gambaran tren dari faktor penyebab perceraian.

Metode analisis *Topic Modeling* yang umum digunakan untuk membuat model topik adalah metode *Latent Dirichlet Allocation* (LDA). Menurut Blei dan Jordan (2003) metode LDA adalah metode *text mining* untuk menemukan pola tertentu pada sebuah dokumen yang tujuannya untuk mendapatkan beberapa macam topik yang berbeda, dengan mengelompokkan dokumen ke dalam sebuah topik tertentu. LDA menghasilkan daftar topik yang diberi bobot untuk masing-masing dokumen. Dengan menggunakan analisis tersebut untuk penelitian ini, maka dapat diketahui beberapa topik yang mendasar dari faktor penyebab perceraian. Sehingga, diharapkan dengan adanya penelitian ini dapat memudahkan dalam memahami dan mengetahui penyebab perceraian yang terjadi di Kota Yogyakarta.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan di atas, maka rumusan masalah yang akan diteliti adalah:

1. Bagaimana gambaran umum pada data putusan perceraian di Pengadilan Agama Yogyakarta pada tahun 2020?
2. Bagaimana hasil implementasi *Topic Modeling* mengenai faktor perceraian dengan metode *Latent Dirichlet Allocation* (LDA) berdasarkan alasan pengajuan perceraian pada dokumen putusan perceraian di Pengadilan Agama Yogyakarta tahun 2020?

1.3. Batasan Masalah

Untuk menghindari meluasnya pembahasan dalam penelitian maka dibutuhkan beberapa batasan masalah yang digunakan yaitu:

1. Data yang digunakan merupakan data sekunder yang diperoleh dari Pengadilan Agama Yogyakarta dan *website* Direktori Putusan Mahkamah Agung Republik Indonesia yaitu <https://putusan3.mahkamahagung.go.id/> diakses dan diambil paling lambat pada tanggal 14 Desember 2021.

2. Jenis putusan yang digunakan adalah putusan perceraian (cerai talak dan cerai gugat) yang berhasil di putus oleh Pengadilan Agama Yogyakarta pada tahun 2020.
3. Data yang digunakan untuk analisis *topic modeling* yaitu teks yang memuat alasan dari pengadu untuk mengajukan perceraian yang termuat dalam bagian duduk perkara dokumen putusan perceraian.
4. Data diolah dengan bantuan *software Python* yang digunakan untuk analisis data teks dan *software Microsoft Excel* untuk menyimpan data dan pengolahan diagram.
5. Metode yang digunakan untuk analisis *topic modeling* adalah *Latent Dirichlet Allocation (LDA)*.

1.4. Jenis Penelitian dan Metode Analisis

Pada penelitian ini, peneliti akan melakukan penelitian untuk Tugas Akhir dengan menggunakan metode penelitian kualitatif. Penelitian Kualitatif merupakan suatu jenis penelitian yang berkaitan dengan analisis data kualitatif, seperti analisis deskriptif, regresi, dan lain sebagainya. Penelitian kualitatif berdasar pada kerangka teori, pemahaman peneliti maupun gagasan ahli yang kemudian dikembangkan menjadi permasalahan dan pemecahan masalah.

Metode analisis data yang digunakan dalam penelitian adalah statistika deskriptif dan pemodelan topik menggunakan metode *Latent Dirichlet Allocation (LDA)* untuk mendapatkan beberapa topik mengenai faktor penyebab perceraian. Adapun tahap analisis dari penelitian adalah analisis deskriptif, *text preprocessing*, pembobotan kata, visualisasi *topics coherence*, dan pemodelan topik. Beberapa *software* yang digunakan untuk membantu analisis dan visualisasi adalah *Microsoft Excel 365* dan *Anaconda3 2021.05 (Python 3.8.8 64-bit)*.

1.5. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dibuat, berikut merupakan tujuan dari penelitian:

1. Untuk mengetahui gambaran umum pada data putusan perceraian di Pengadilan Agama Yogyakarta pada tahun 2020.

2. Untuk mengetahui hasil implementasi *Topic Modeling* mengenai faktor perceraian dengan metode *Latent Dirichlet Allocation* (LDA) berdasarkan alasan pengajuan perceraian pada dokumen putusan perceraian di Pengadilan Agama Yogyakarta tahun 2020.

BAB II

TINJAUAN PUSTAKA

Peneliti menggunakan penelitian terdahulu sebagai acuan dalam penulisan penelitian yang bertujuan untuk mengetahui hubungan antara penelitian terdahulu dengan penelitian yang dilakukan saat ini. Hasil dari penelitian terdahulu diharapkan dapat dikembangkan dan diterapkan pada teori yang hendak diteliti. Pada penelitian ini berkaitan dengan *topic modeling* pada teks dokumen. Berikut beberapa penelitian yang digunakan peneliti sebagai acuan dalam penelitian.

Penelitian pertama yang dilakukan oleh Merchant dan Pande (2018) berjudul “*NLP Based Latent Semantic Analysis for Legal Text Summarization*”. Penelitian ini berkaitan dengan *topic modeling* dengan menggunakan metode *Latent Semantic Analysis* (LSA) dengan tujuan meringkas teks dokumen pada kasus-kasus di pengadilan. Ringkasan singkat dibuat dengan tetap menjaga topik penting dari dokumen asli. Didapatkan rata-rata skor ROGUE-1 sebesar 0,58. Namun, evaluasi dari metode ini tidak sepenuhnya efektif dan diharapkan pada penelitian selanjutnya dapat mengevaluasi model tidak hanya pada kata-kata yang mirip tetapi juga berdasarkan seluruh konsep. Menggunakan metode LSA terkadang menyebabkan jeda dalam kontinuitas keseluruhan.

Penelitian kedua dilakukan oleh Annisa, Surjandari, dan Zulkarnain (2019) yang berjudul “*Opinion Mining on Mandalika Hotel Reviews Using Latent Dirichlet Allocation*”. Penelitian ini bertujuan untuk menyimpulkan topik-topik yang berasal dari ulasan hotel Mandalika di Provinsi Nusa Tenggara Barat menggunakan metode *Latent Dirichlet Allocation* (LDA). Hasil dari penelitian yaitu diperoleh delapan topik yang sering dibahas. Jumlah topik ditentukan berdasarkan nilai *coherence* sebagai ukuran evaluasi. Penelitian ini diharapkan dapat menjadi masukan untuk manajemen hotel dan wisatawan dalam rangka meningkatkan jumlah kunjungan wisatawan ke Mandalika.

Penelitian ketiga berjudul “*Analisis Metoda Latent Dirichlet Allocation untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik*” yang dilakukan oleh Setijohatmo, dkk (2020). Penelitian ini berkaitan dengan pemodelan topik dengan menggunakan metode *Latent Dirichlet Allocation* (LDA)

pada dokumen laporan tugas akhir Jurusan Teknik Komputer. Penelitian menggunakan sekumpulan laporan tugas akhir yang telah diberi label. Hasil penelitian diukur tingkat korelevanannya jika dibandingkan dengan *judgement* manusia dalam bentuk laporan tugas akhir berlabel. Dari penelitian ini diperoleh bahwa apabila teks yang dianalisis mengandung banyak kata umum maka akan mengurangi tingkat presisi pada topik yang dihasilkan. Selain itu diperoleh bahwa metode LDA tidak menghasilkan topik yang berlabel dalam mengelompokkan teks dokumen.

Penelitian keempat berjudul “*Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter*” yang dilakukan oleh Xue, dkk (2020). Penelitian ini bertujuan untuk memahami wacana dan reaksi psikologis pengguna Twitter terhadap COVID-19 dengan menggunakan metode *Latent Dirichlet Allocation* (LDA). Diperoleh sebelas topik yang menonjol diidentifikasi dan kemudian dikategorikan ke dalam sepuluh tema. Hasil topik tidak mengungkapkan pengobatan dan gejala sebagai topik umum di Twitter. Analisis sentimen menunjukkan bahwa ketakutan akan hal yang tidak diketahui terhadap virus corona dominan di semua topik.

Penelitian kelima dilakukan oleh Sanjaya (2021) yang berjudul “Implementasi *Latent Dirichlet Allocation* (LDA) untuk Klasterisasi Cerita Berbahasa Bali”. Tujuan dari penelitian ini adalah untuk mengetahui topik-topik yang tersembunyi pada kumpulan cerita berbahasa Bali. Jumlah kelompok ditentukan berdasarkan masukan yang disesuaikan dengan jumlah kelompok yang dibuat oleh ahli bahasa Bali yaitu sebanyak 3 kelompok. Diperoleh hasil pengelompokan dengan akurasi tertinggi adalah 62% pada jumlah kata yang digunakan sebagai representasi dokumen berjumlah 3000 kata. Selain itu, didapatkan kesimpulan bahwa akurasi pengelompokan dipengaruhi oleh ukuran kesamaan yang digunakan ketika melakukan penggabungan dokumen serta jumlah kata sebagai representasi dokumen.

Penelitian keenam dilakukan oleh Rashif, dkk (2021) yang berjudul “Implementasi LDA untuk Pengelompokan Topik Cuitan Akun Bot Twitter bertagar #Covid-19”. Penelitian ini bertujuan untuk mengetahui topik-topik yang sering dibahas terhadap penyebaran informasi menggunakan tagar #covid19.

Hasil dari penelitian diperoleh bahwa terdapat lima topik teratas yang berkaitan dengan kondisi dan dampak pandemi saat ini, himbauan untuk selalu menjaga jarak, perkembangan kasus covid-19 di Indonesia, vaksinasi di beberapa wilayah Indonesia, dan cara menghadapi pandemi covid-19.

Penelitian ketujuh dilakukan oleh Pathana dan Prakash (2021) yang berjudul “*Unsupervised Aspect Extraction Algorithm for opinion mining using topic modeling*”. Penelitian ini berkaitan dengan memodelkan topik dengan menggunakan tiga metode populer yaitu *Latent Dirichlet Allocation* (LDA), *Latent Semantic Analysis* (LSA) dan *Hierarchical Dirichlet Process* (HDP) pada beberapa ulasan seperti ulasan alat musik, ulasan ponsel, ulasan otomotif, ulasan restoran dan database film internet (IMDb). Pada penelitian menunjukkan hasil bahwa ketiga metode bekerja dengan baik untuk proses pemodelan topik, tetapi metode LDA memiliki nilai *coherence* yang lebih baik daripada LSA dan HDP. Sehingga diperoleh hasil evaluasi bahwa metode LDA adalah model terbaik untuk memodelkan topik pada beberapa ulasan publik.

Tabel 2.1 Tabel Penelitian Sebelumnya

No	Penulis	Judul	Metode	Persamaan	Perbedaan
1	Kaiz Merchant dan Yash Pande (2018)	<i>NLP Based Latent Semantic Analysis for Legal Text Summarization</i>	<i>Latent Semantic Analysis</i> (LSA)	Studi kasus sama-sama menggunakan dokumen hukum	Metode yang digunakan berbeda
2	Rossi Annisa, Isti Surjandari, dan Zulkarnain (2019)	<i>Opinion Mining on Mandalika Hotel Reviews Using Latent Dirichlet Allocation</i>	<i>Latent Dirichlet Allocation</i> (LDA)	Metode yang digunakan sama	Studi kasus yang digunakan berbeda
3	Urip T. Setijohatmo, Setiadi Rachmat, Tati Susilawati, Yuda Rahman (2020)	Analisis Metoda <i>Latent Dirichlet Allocation</i> untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik	<i>Latent Dirichlet Allocation</i> (LDA)	Metode yang digunakan sama	Studi kasus yang digunakan berbeda
4	Jia Xue, Junxiang Chen, Chen Chengda	<i>Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet</i>	<i>Latent Dirichlet Allocation</i> (LDA) untuk analisis	Metode yang digunakan sama	Studi kasus yang digunakan berbeda

No	Penulis	Judul	Metode	Persamaan	Perbedaan
	Zheng, Sijia Li, dan Tingshao Zhu (2020)	<i>Allocation for topic modeling on Twitter</i>	sentimen		
5	Ngurah Agus Sanjaya ER (2021)	Implementasi <i>Latent Dirichlet Allocation</i> (LDA) untuk Klasterisasi Cerita Berbahasa Bali	<i>Latent Dirichlet Allocation</i> (LDA) untuk klasterisasi	Metode yang digunakan sama	Studi kasus yang digunakan berbeda
6	Faza Rashif, Goldio Ihza Perwira Nirvana, Muhammad Alif Noor, dan Nur Aini Rakhm (2021)	Implementasi LDA untuk Pengelompokan Topik Cuitan Akun Bot Twitter bertagar #Covid-19	<i>Latent Dirichlet Allocation</i> (LDA)	Metode yang digunakan sama	Studi kasus yang digunakan berbeda
7	Azizkhan F Pathana dan Chetana Prakash (2021)	<i>Unsupervised Aspect Extraction Algorithm for opinion mining using topic modeling</i>	<i>Latent Dirichlet Allocation</i> (LDA), <i>Latent Semantic Analysis</i> (LSA) dan <i>Hierarchical Dirichlet Process</i> (HDP)	Salah satu metode yang digunakan sama	Studi kasus yang digunakan berbeda

BAB III

LANDASAN TEORI

3.1. Perceraian

Menurut Kamus Besar Bahasa Indonesia istilah “carai” adalah kata kerja yang artinya 1. Pisah, 2. Putus sebagai istri; talak. Begitu juga istilah “cerai” adalah kata benda yang artinya: 1. Perpisahan, 2. Perihal bercerai (antara suami dan istri), berpisah. Adapun istilah “bercerai” adalah kata kerja yang artinya 1. Tidak bercampur (hubungan, bersatu, dsb) 2. Berarti berlaki-bini (suami-istri).

Perceraian adalah peristiwa di mana suami-istri berpisah dalam satu kesatuan keluarga karena perbedaan prinsip yang tidak dapat dipersatukan kembali dengan berbagai cara. Perceraian diatur secara tegas pada Pasal 39 ayat 1 dan 2 Undang-Undang tentang Perkawinan Nomor 1 Tahun 1974 yang berbunyi sebagai berikut: 1). Perceraian hanya dapat dilakukan di depan sidang pengadilan yang bersangkutan telah berusaha dan tidak berhasil untuk mendamaikan kedua belah pihak. 2) Perceraian harus terdapat alasan yang cukup antara suami dan istri tidak akan bisa akur sebagai suami istri.

Hukum Islam membenarkan dan menghalalkan perceraian bila perceraian lebih membawa kepada keadaan yang lebih baik daripada tetap dalam ikatan perkawinan yang di dalamnya terdapat siksaan batin. Walaupun maksud berasal perkawinan artinya buat mencapai kebahagiaan serta kerukunan hati masing-masing, tentulah kebahagiaan tidak dapat tercapai dalam hal-hal yang tak dapat disesuaikan sebab intinya kebahagiaan tidak dapat dipaksakan. Memaksa kebahagiaan bukanlah kebahagiaan yang datang dari diri sendiri, sehingga dengan sendirinya akan menjadi beban yang berujung pada penderitaan, oleh karena itu Islam tidak mengikat perkawinan, tetapi tidak pula mempermudah pernikahan (Latief, 1985).

3.2. Alasan Perceraian

Berikut beberapa alasan tertentu mengenai faktor penyebab perceraian yang mungkin terjadi:

- a. Zina;
- b. Meninggalkan tempat tinggal bersama;

- c. Dijatuhkan hukuman penjara lima tahun atau hukuman yang lebih berat lagi, setelah dilangsungkan perkawinan.
- d. Pencederaan berat atau penganiayaan yang dilakukan oleh salah seorang diri suami istri tersebut terhadap yang lain, sehingga dapat menyebabkan luka yang berbahaya bahkan membahayakan keselamatan jiwa (Soimin, 2016).

Alasan perceraian juga diatur dalam Pasal 39 UU No 1 Tahun 1974 dan juga dalam Pasal 19 peraturan pemerintah No 9 tahun 1975 tentang pelaksanaan UU No 1 Tahun 1974, yaitu:

- a. Salah satu pihak berbuat zina menjadi pemabuk, pematik, penjudi, dan lain sebagainya yang sukar disembuhkan;
- b. Salah satu pihak meninggalkan pihak lain selama dua tahun berturut-turut tanpa izin pihak lain dan tanpa alasan yang sah karena hal lain di luar kemampuannya;
- c. Salah satu pihak mendapat hukuman penjara lima tahun atau hukuman yang lebih berat setelah perkawinan berlangsung;
- d. Salah satu pihak melakukan kekejaman atau penganiayaan berat yang membahayakan pihak lain;
- e. Salah satu pihak mendapat cacat badan atau penyakit dengan akibat tidak dapat menjalankan kewajiban sebagai suami/istri;
- f. Antara suami dan istri terus menerus terjadi perselisihan dan pertengkaran dan tidak ada harapan akan hidup rukun lagi dalam rumah tangga (Sudarsono, 2005)

3.3. Macam-Macam Perceraian

Berdasarkan ketentuan mengenai perceraian yang diatur dalam Undang-Undang No 1 Tahun 1974 hingga peraturan pemerintah, kasus perceraian dibedakan menjadi dua macam yaitu cerai talak dan cerai gugat. Perceraian disebabkan adanya permohonan seorang suami untuk menceraikan istrinya yang disebut cerai talak dan gugatan seorang istri yang ingin bercerai dengan suaminya disebut dengan cerai gugat, berikut merupakan penjelasan lebih lanjut.

- a. Cerai Talak

Cerai talak adalah apabila suami yang mengajukan permohonan ke pengadilan untuk menceraikan istrinya, dan istri tersebut menyetujuinya (Ali, 2006). Secara sederhana talak ialah perbuatan yang menyebabkan putusnya ikatan perkawinan yang dengan itu pula gugurlah kehalalan hubungan antara suami istri.

Di dalam pasal 66 Undang-Undang No. 7 Tahun 1989 tentang Peradilan Agama menyatakan:

1. Seorang suami yang beragama Islam yang akan menceraikan istrinya mengajukan permohonan kepada pengadilan untuk mengadakan sidang guna menyaksikan ikrar talak.
2. Permohonan sebagaimana yang dimaksud dalam ayat (1) diajukan kepada pengadilan yang daerah hukumnya meliputi tempat kediaman termohon kecuali apabila termohon dengan sengaja meninggalkan tempat kediaman yang ditentukan bersama tanpa izin pemohon.
3. Dalam hal termohon bertempat kediaman di luar negeri permohonan diajukan kepada pengadilan yang daerah hukumnya meliputi tempat kediaman pemohon.
4. Dalam hal pemohon dan termohon bertempat kediaman di luar negeri, maka permohonan diajukan kepada pengadilan yang daerah hukumnya meliputi tempat perkawinan mereka dilangsungkan atau kepada Pengadilan Agama Jakarta Pusat.
5. Permohonan soal penguasaan anak, nafkah anak, nafkah istri dan harta bersama suami istri dapat diajukan bersama-sama dengan permohonan cerai talak ataupun sesudah ikrar talak diucapkan.

Cerai talak dibedakan menjadi dua macam yaitu sebagai berikut:

- Talak Sunni

Talak sunni merupakan talak yang diberikan oleh suami kepada istrinya dalam bentuk murni dan yang belum melakukan hubungan seksual selama waktu suci. Dalam hal ini termasuk talak raj'i dan talak ba'in. Talak raj'i adalah talak yang diberikan oleh suami kepada istri yang telah digauli, yang selama masa 'iddah dapat suami dapat kembali rujuk dengan istri tanpa adanya persetujuan dari istri. Talak ba'in dibagi

menjadi dua jenis, yaitu talak ba'in shugra adalah talak yang menghilangkan mantan suami untuk kembali istri selama masa 'iddah atau sesudahnya, kecuali dengan akad nikah baru dan maskawin. Dan talak ba'in kubra adalah talak yang menghilangkan mantan suami untuk kembali istri pada saat iddah atau sesudahnya, tetapi wanita tersebut harus menikah dengan pria lain dan bercerai dalam arti kata yang sebenarnya.

- Talak Bid'i

Talak Bid'i merupakan talak yang diberikan suami kepada istrinya yang bertentangan dengan syara'. Talak yang termasuk talak bid'i adalah talak tiga yang diberikan suami kepada istri ketika dia sedang haid atau melahirkan (Ks, 2009).

b. Cerai Gugat

Sebagaimana diuraikan di atas, perkara cerai gugat merupakan perceraian yang diajukan oleh istri ke pengadilan karena putusnya perkawinan dengan suaminya. Dalam Peraturan Pemerintah No 9 tahun 1975 pasal 19 tentang perkawinan menurut agama Islam dapat berupa gugatan karena suami melanggar taklik talak, gugatan karena syiqoq, gugatan karena fasakh, dan gugatan karena alasan-alasan lain.

3.4. Statistika Deskriptif

Secara garis besar statistik dibagi menjadi dua yaitu statistik deskriptif dan statistik inferensi. Statistik deskriptif merupakan rangkuman yang mengandung ukuran-ukuran yang menjadi karakter dari suatu data (Walpole, Myers, Myers, & Ye, 2011). Metode deskriptif adalah penulisan yang dilakukan untuk mengetahui keberadaan variabel mandiri, baik hanya pada satu variabel atau lebih (variabel yang berdiri sendiri) tanpa membuat perbandingan dan mencari hubungan variabel itu dengan variabel yang lain (Sugiyono, 2009). Statistika deskriptif adalah bagian dari ilmu statistika yang hanya mengolah, menyajikan data dengan hanya melihat gambaran secara umum dari data yang didapatkan. Ciri - ciri analisis data deskriptif yaitu penyajian data lebih ditekankan dalam bentuk tabel, grafik, dan ukuran-ukuran statistik seperti persentase, rata-rata, variansi, korelasi, dan angka indeks. Beberapa hal yang dapat dilakukan adalah penyajian data melalui tabel,

grafik, diagram lingkaran, dan lain-lain. Selain itu, analisis ini tidak menggunakan uji signifikansi dan taraf kesalahan karena tidak ada kesalahan generalisasi.

3.5. Artificial Intelligence (AI)

Menurut Wijaya (2013) *Artificial Intelligence* (AI) atau kecerdasan buatan adalah bagian dari ilmu komputer yang menjadikan komputer (mesin) menjadi cerdas sehingga dapat menggantikan atau meniru pekerjaan yang dilakukan oleh manusia. Pada awalnya komputer hanya dibuat sebagai alat komputasi, namun dengan berkembangnya zaman komputer diharapkan dapat digunakan untuk melakukan pekerjaan yang dilakukan oleh manusia. Untuk membuat komputer menjadi cerdas, perlu ditambahkan komponen pengetahuan dan keterampilan penalaran ke komputer. Berikut adalah dua bagian utama yang diperlukan untuk menerapkan konsep kecerdasan buatan ke komputer (Andayu, 2013).

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

Menurut Andayu (2013), tujuan utama AI adalah sebagai berikut.

1. Sistem pakar (*Expert System*), komputer digunakan sebagai media penyimpanan pengetahuan pakar agar komputer memiliki pengalaman memecahkan masalah dengan meniru keahlian 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 suaranya.
4. Robotika Sistem Sensor (*Robotics and Sensory Systems*), contoh bagian robotika dari sistem sensor seperti sistem penglihatan dan pencitraan dan sistem pemrosesan sinyal.
5. *Computer Vision*, untuk menerjemahkan suatu gambar ke dalam komputer.
6. *Intelligent Computer-aided Instruction*, komputer digunakan sebagai tutor yang dapat melatih dan mengajar.

7. Mesin belajar (*Machine Learning*), adalah sekumpulan metode pemecahan masalah yang digunakan untuk melatih komputer dengan menganalisis kasus-kasus yang telah terjadi.

3.6. Text Mining

Text mining adalah istilah untuk mengekstrak informasi berupa teks yang datanya bersumber dari dokumen. Menurut Feldman, dkk (2007) tujuan dari *text mining* adalah menemukan kata-kata yang dapat mewakili isi dokumen sehingga dapat dicari hubungan antar dokumen satu dengan yang lainnya (Hikmah, 2018). Penambangan teks adalah teknik yang digunakan untuk mengelola pengelompokan, klasifikasi, ekstraksi informasi, dan pengambilan informasi (Hudaya, Fakhurroja, & Alamsyah, 2019). Analisis data yang mengandung bahasa alami digunakan teknik dan alat untuk merancang, menemukan, dan mengekstrak pengetahuan menjadi data dengan argumen yang lebih terstruktur dan dapat dianalisis lebih mudah. Text mining bekerja pada data teks yang tidak terstruktur dan diharapkan dapat mengubah teks menjadi lebih terstruktur dan dapat dengan mudah dianalisis dengan mengubah kata atau kalimat (Kabiru & Sari, 2019). Dalam proses eksplorasi teks, ada empat tahap, yaitu, pemrosesan awal terhadap teks (*text preprocessing*), transformasi teks (*text transformation*), pemilihan fitur (*feature selection*), dan penemuan pola (*pattern discovery*) (Sanjaya & Absar, 2015).

3.7. Text PreProcessing

Text PreProcessing merupakan tahap awal pada analisis teks dokumen dengan tujuan untuk mengubah data teks menjadi lebih terstruktur. Hasil yang diperoleh pada tahap *text preprocessing* adalah berupa nilai numerik sehingga dapat dianalisis menggunakan perhitungan lebih lanjut (Socrates, Akbar, & Akbar, 2016). Pada tahap *text preprocessing* juga bertujuan untuk mengubah teks menjadi sederhana sehingga analisis dapat bekerja dengan baik. *Text preprocessing* merupakan tahap menghilangkan karakter-karakter yang tidak diperlukam pada dokumen (Siregar, Sinaga, & Arianto, 2017).

Tahap *pre-processing* merupakan tahap yang berpengaruh dan dapat memberikan hasil setelah mengolah data teks. Jika tahap ini dilakukan dengan benar, maka hasil dari langkah pre-processing juga akan mendapatkan hasil yang

baik. Proses *preprocessing* terdiri dari beberapa proses yaitu *lowercasing*, *remove punctuation*, *stopword removal*, *lemmatization* dan *tokenizing*.

1. Lowercasing

Tahap awal *preprocessing* merupakan tahap dimana teks akan diubah dari huruf besar menjadi huruf kecil, sehingga karakter yang akan diterima hanya dari “a” menjadi “z” (Socrates, Akbar, & Akbar, 2016). Dalam hal ini untuk menghindari adanya dua kata yang identik tetapi dianggap berbeda dari program karena perbedaan huruf besar dan kecil (Hadi, W., & Hasan, 2017).

2. Remove Punctuation

Remove punctuation merupakan tahaps yang dilakukan untuk menghilangkan karakter yang tidak terpakai berupa tanda baca, angka, markup/html/tag, karakter khusus (\$, %, -, dst) (Cendana & Permana, 2019). Karakter-karakter tersebut memungkinkan mengganggu dalam pemrosesan analisis teks (Hadi, W., & Hasan, 2017).

3. Stopword Removal

Stopwords adalah kata-kata yang terdapat pada dokumen yang tidak memiliki makna sehingga dapat dihilangkan dalam analisis (Socrates, Akbar, & Akbar, 2016). Sedangkan *stopword removal* merupakan tahap menghilangkan kata-kata yang tidak memiliki makna menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting).

4. Lemmatization

Menurut Ingason, dkk (2008), *lemmatization* adalah sebuah proses untuk menemukan bentuk dasar dari sebuah kata. Nirenburg (2009) menjelaskan bahwa *lemmatization* adalah proses yang bertujuan untuk melakukan normalisasi pada teks/kata dengan berdasarkan pada bentuk dasar yang merupakan bentuk lemanya. Normalisasi diartikan dengan mengidentifikasi dan menghapus prefiks serta sufiks dari sebuah kata. Lema adalah bentuk dasar dari sebuah kata yang memiliki arti tertentu berdasar pada kamus.

5. Tokenizing

Tokenizing merupakan tahap pemisahan atau pemotongan teks menurut setiap kata yang menyusunnya (Socrates, Akbar, & Akbar, 2016). Pemisahan kata ini bertujuan untuk memudahkan proses selanjutnya dalam penghitungan kata,

pembobotan kata sampai diubah menjadi vektor berdimensi tinggi (Hadi, W., & Hasan, 2017).

3.8. Pembobotan Kata (*Term Weighting*)

Pembobotan kata (*term weighting*) merupakan suatu proses pembobotan pada kata. Pembobotan dasar dilakukan dengan menghitung frekuensi kemunculan *term* (kata) dalam dokumen. Frekuensi kemunculan (*term frequency*) merupakan petunjuk seberapa baik term mewakili isi dokumen. Semakin tinggi kemunculan suatu *term* dalam dokumen maka akan semakin tinggi pula nilai kesesuaian. Ketika frekuensi kata lebih sering muncul dalam dokumen tertentu atau kejarangmunculan kata (*term scarcity*), maka menunjukkan bahwa bobot meningkat dan kata tersebut dianggap penting (*uncommon terms*). Jika kata tersebut sering muncul di banyak dokumen, ini menunjukkan bahwa bobot berkurang dan seberapa umum kata tersebut. Pembobotan akan memperhitungkan faktor kebalikan dari frekuensi dokumen yang mengandung suatu kata (*inverse document frequency*) (Karmayasa & Mahendra, 2012).

3.8.1 *Term Frequency (TF)*

Term Frequency (TF) adalah suatu metode pembobotan *heuristic* yang menentukan bobot dokumen berdasarkan kemunculan *term*. Hasil dari pembobotan kata yang telah dilakukan selanjutnya digunakan untuk menentukan dokumen yang relevan. Terdapat beberapa cara untuk menghitung nilai TF, diantaranya sebagai berikut (Yoren, 2018):

1. *Raw Term Frequency (TF Murni)*, dengan menghitung frekuensi kemunculan suatu term pada dokumen. Misalkan muncul term sebanyak 3 kali maka *term* tersebut akan bernilai 3.
2. *Binary Term Frequency (TF Binary)*, menyeragamkan bobot pada dokumen terhadap term dengan nilai 1 dan 0. Nilai 1 jika muncul minimal satu kali dalam dokumen, nilai 0 jika kata tidak muncul sama sekali.
3. *Logarithmic Term Frequency (TF Logaritmik)*, dengan menggunakan fungsi logaritmik yang dimaksudkan untuk mengurangi dominasi dokumen yang mengandung sedikit *term* dalam *query* namun mempunyai frekuensi yang tinggi. Nilai TF Logaritmik yang dapat diperoleh dari persamaan berikut:

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

4. *Augmented Term Frequency* (TF Normalisasi), meyeragamkan bobot dokumen terhadap *term* dengan memberikan range 0.5 sampai 1 dengan membandingkan antara frekuensi *term* dengan jumlah *term* terbanyak yang muncul pada dokumen yang sama. Nilai TF Normalisasi yang dapat diperoleh dari persamaan berikut:

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

3.8.2 *Inverse Document Frequency (IDF)*

Inverse Document Frequency (IDF) adalah nilai kepentingan suatu dokumen terhadap *term*. IDF dapat menunjukkan hubungan *term* dalam sebuah dokumen. Misalkan term ‘adalah’ sering muncul dalam berbagai dokumen sehingga *term* ini bukan kata kunci yang baik untuk membedakan dokumen yang relevan. Sehingga, IDF mengurangi bobot istilah yang sangat sering muncul dalam dokumen (Yoren, 2018). Nilai IDF dapat dihitung dengan menggunakan persamaan sebagai berikut:

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

dengan:

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

D : Jumlah semua dokumen

DF_t : Jumlah dokumen yang mengandung *term* (t)

3.8.3 *Term Frequency-Inverse Document Frequency (TF-IDF)*

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) adalah suatu metode untuk memberikan bobot pada suatu *term* dalam suatu dokumen yang sering digunakan dalam *text mining*. Metode TF-IDF merupakan gabungan dari dua konsep perhitungan bobot, yaitu kombinasi *term frequency* (tf) / frekuensi kemunculan suatu kata dalam sebuah dokumen dan *inverse document frequency* (idf) / invers dari frekuensi dokumen yang mengandung kata (Karmayasa & Mahendra, 2012). Metode TF-IDF dilakukan dengan mengalikan dua metrik dengan menghitung bobot yaitu berapa kali pesan muncul kata dalam dokumen dan invers frekuensi dokumen yang berisi kata (Nurjannah, Hamdani, & Astuti, 2013). Nilai TF-IDF dapat dihitung dengan menggunakan persamaan sebagai berikut:

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

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

dengan:

d : dokumen ke- d

t : *term* ke- t

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

TF_{dt} : Jumlah kemunculan term (t) dalam dokumen (d)

Berdasarkan persamaan 3.5 apabila nilai $D = DF_t$ maka akan menghasilkan nilai 0 karena hasil dari perhitungan $\ln(1) = 0$ sehingga seberapa besar nilai TF_{dt} tidak akan berpengaruh dalam perhitungan bobot. Sehingga perlu ditambahkan nilai 1 pada sisi IDF, sehingga menghasilkan persamaan sebagai berikut:

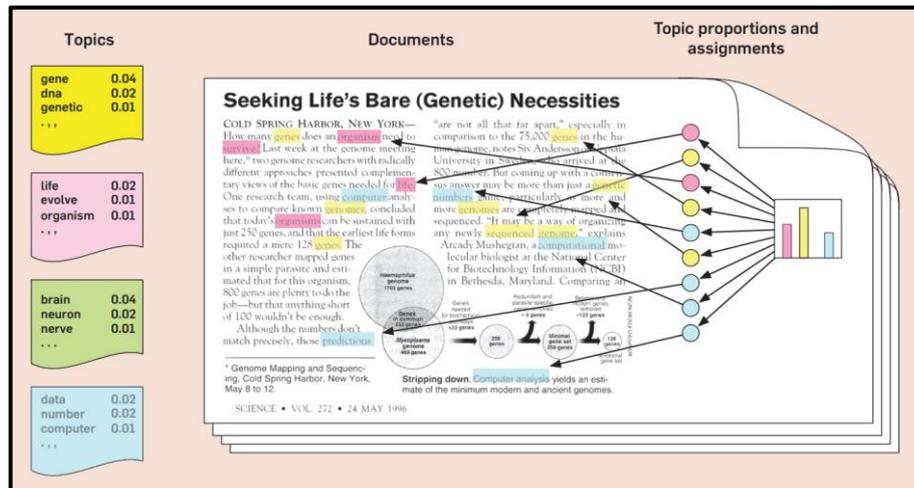
$$W_{dt} = TF_{dt} \times \left(\ln \left(\frac{D}{DF_t} \right) + 1 \right) \quad (3.6)$$

3.9. *Topic Modeling*

Konsep *topic modeling* menurut Blei (2003) terdiri dari beberapa entitas yaitu “kata-kata (*words*)”, “dokumen (*documents*)”, dan “korpora (*corpora*)”. Sebuah kata merupakan unit dasar dari data, didefinisikan sebagai item dari kosakata yang diberi indeks untuk kata tertentu yang ada pada dokumen. Dokumen merupakan urutan kata-kata. Korpus (*corpus*) merupakan kumpulan dokumen dan korpora merupakan bentuk jamak dari korpus. Topik (*topic*) merupakan distribusi dari beberapa kosakata yang tetap. Secara sederhana, setiap dokumen dalam korpus berisi proporsi tersendiri dari beberapa topik yang dibahas sesuai dengan kata yang terkandung dalam dokumen. *Topic modeling* merupakan pemodelan yang terdiri dari beberapa kata tertentu yang menyusun topik, dan dalam suatu dokumen memiliki probabilitas tertentu dari beberapa topik yang dihasilkan.

Topic modeling merupakan metode yang digunakan untuk menemukan topik utama yang tersembunyi dari rangkaian kata dalam kumpulan dokumen yang besar dan tidak terstruktur. Metode *topic modeling* menganalisis data berdasarkan teks asli, mengenai hubungan antar topik satu sama lain, hubungan antar tema yang dapat berubah sewaktu-waktu, sehingga metode ini dapat dikembangkan untuk pencarian atau meringkas teks yang terdapat dalam dokumen. Secara umum, setiap dokumen dalam korpus akan memiliki proporsi topik yang dibahas. Hasil analisis pemodelan tematik akan digunakan untuk merangkum,

memvisualisasikan, mengeksplorasi dan berteori tentang korpus. Sehingga, *topic modeling* bertujuan untuk menemukan kumpulan topik dan kata yang terkandung dalam berbagai suatu dokumen (Putra & Kusumawardani, 2017). Ilustrasi *topic modeling* menurut Blei (2012), dapat dilihat pada **Gambar 3.1** berikut.



Gambar 3.1 Ilustrasi *Topic Modeling*

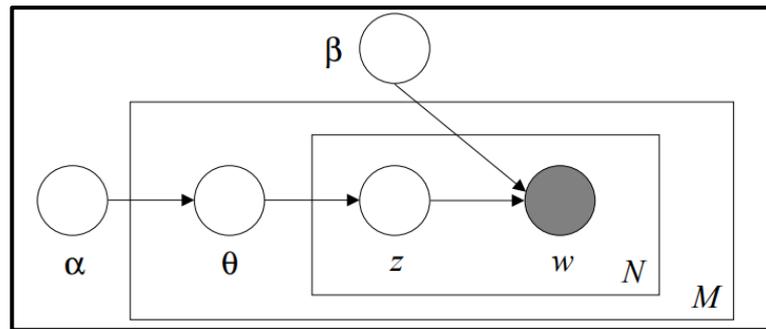
Sumber: (Blei, 2012)

Blei (2012) mengilustrasikan *topic modeling* seperti pada **Gambar 3.1**. Salah satu cara untuk memikirkan tentang bagaimana proses pemodelan topik bekerja adalah dengan membayangkan mengerjakan sebuah artikel dengan sekumpulan *highlighters*. Untuk menemukan kata kunci dapat menggunakan beberapa warna yang membedakan beberapa topik. Sehingga didapatkan beberapa kata yang dikelompokkan berdasarkan warna yang ditetapkan sehingga menjadi sebuah topik (Brett, 2012).

3.10. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) merupakan salah satu metode topik modeling dan topik analisis yang paling populer dan sering digunakan untuk dokumen yang berukuran besar tidak terstruktur. LDA dapat digunakan untuk meringkas, mengelompokkan, menghubungkan, dan memproses data karena menghasilkan beberapa topik yang diberi bobot pada masing-masing dokumen. LDA merupakan algoritma yang didasarkan pada *generative probabilistic model* dengan asumsi setiap topik merupakan campuran dari kumpulan kata-kata dan setiap dokumen merupakan campuran dari berbagai macam topik. Distribusi yang digunakan untuk mendapatkan topik pada dokumen disebut distribusi *Dirichlet*,

selain itu hasil dari *Dirichlet* digunakan untuk mengalokasikan kata-kata ke topik yang berbeda. Dalam LDA, dokumen merupakan objek yang diamati, sedangkan topik, distribusi topik, penggolongan setiap kata pada topik merupakan struktur yang tersembunyi, sehingga metode ini disebut *Latent Dirichlet Allocation* (LDA) (Blei, 2012). Berikut ilustrasi metode LDA secara visual menurut Blei (2003) seperti pada **Gambar 3.2** berikut.



Gambar 3.2 Ilustrasi Metode LDA

Sumber: (Blei, Ng, & Jordan, 2003)

Parameter α merupakan parameter *dirichlet distribution* yang berguna untuk mengontrol distribusi topik pada setiap dokumen, dimana apabila nilai α yang semakin tinggi menyatakan bahwa dokumen memuat topik campuran atau memuat beberapa topik. Sedangkan nilai α yang semakin rendah menyatakan bahwa tidak terdapat topik yang tercampur satu sama lain dalam dokumen. Oleh karena itu, untuk melihat seberapa baik sebaran topik yang dihasilkan dapat dilihat dari nilai α yang semakin rendah. Parameter β merupakan parameter *dirichlet distribution* yang digunakan untuk mengontrol distribusi kata pada setiap topik, dimana apabila nilai β yang semakin tinggi menyatakan bahwa topik memuat kata-kata yang terdapat pada topik lainnya. Sedangkan nilai β yang rendah menyatakan bahwa sebuah topik terdiri dari kata-kata yang lebih spesifik. Variabel θ merupakan distribusi multinomial atau distribusi topik untuk dokumen tertentu, yang menyatakan probabilitas dokumen tertentu. Apabila semakin tinggi nilai θ maka menyatakan bahwa semakin banyak topik yang termuat dalam dokumen, sedangkan semakin rendah nilai θ maka menyatakan bahwa dokumen semakin spesifik pada topik tertentu. Variabel Z merepresentasikan topik untuk kata tertentu dari suatu dokumen. Variabel W merepresentasikan kata spesifik yang berkaitan dengan topik tertentu dalam suatu dokumen. Bentuk lingkaran

menyatakan bahwa kata bersifat individual. Variabel W digambarkan dengan lingkaran yang berwarna abu-abu merepresentasikan variabel yang diobservasi, sedangkan lingkaran yang berwarna putih merepresentasikan variabel laten atau atau *hidden* variabel yang secara tidak langsung diobservasi (Blei, Ng, & Jordan, 2003). Berikut adalah perhitungan probabilitas dari sebuah *corpus* berdasarkan notasi yang telah disebutkan.

$$p(D|\alpha, \beta) = \prod_{d=1}^M \int p(\theta_d|\alpha) \left(\prod_{n=1}^{N_d} \sum_{Z_{dn}} p(Z_{dn}|\theta_d) p(w_{dn}|Z_{dn}, \beta) \right) d\theta_d \quad (3.7)$$

dengan:

p : Nilai probabilitas

M : Banyaknya dokumen

N : Banyaknya kata dalam suatu dokumen

α : Nilai distribusi topik pada setiap dokumen

β : Nilai distribusi kata pada setiap topik

θ_d : Vektor probabilitas distribusi topik dalam suatu dokumen

Z_{dn} : Vektor identitas topik untuk kata tertentu dari suatu dokumen

W_{dn} : Vektor identitas kata tertentu dalam suatu dokumen

Secara umum metode LDA bekerja dengan memasukkan data dan beberapa tolak ukur yang akan menghasilkan suatu model yang tersusun dari beberapa bobot yang telah dinormalisasi dengan nilai probabilitas. Nilai probabilitas mengacu pada dua jenis, yaitu (1) jenis probabilitas suatu dokumen spesifik tertentu menghasilkan topik yang spesifik pula dan (2) jenis probabilitas suatu topik spesifik tertentu yang menghasilkan kata-kata spesifik dari kumpulan kosakata. Probabilitas (1) dari dokumen yang sudah diberi label dengan beberapa topik biasanya dilanjutkan hingga menghasilkan probabilitas (2) yang menghasilkan kata-kata spesifik (Campbell, Hindle, & Stroulia, 2014).

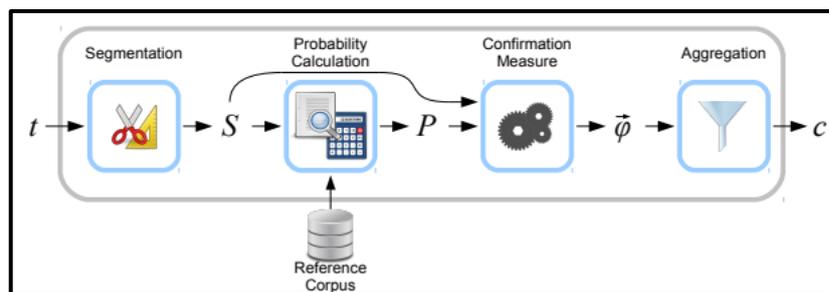
3.11. *Topic Coherence*

Topic coherence merupakan nilai yang digunakan untuk mengevaluasi pemodelan topik. Ukuran *topic coherence* menilai suatu topik dengan mengukur tingkat kesamaan semantik antara kata-kata dengan skor tinggi dalam topik tersebut. Pengukuran ini membantu membedakan antara topik yang merupakan topik yang dapat ditafsirkan secara semantik dan topik yang merupakan artefak dari statistik inferensi (Stevens, Kegelmeyer, Andrzejewski, & Buttler, 2012).

Topic coherence dianggap lebih mampu memberikan interpretasi yang lebih baik terhadap hasil *topic modeling* dibandingkan dengan *Perplexity* (Lestari, 2019).

Pembagian jumlah topik pada proses LDA merupakan hal yang penting untuk dilakukan dalam proses evaluasi pemodelan topik. Hasil dari penentuan jumlah topik dapat membantu dalam menunjukkan optimal dan validnya model tersebut agar dapat dipahami oleh manusia. Teknik evaluasi untuk pemodelan topik, salah satunya adalah *Coherence score*. *Coherence score* merupakan acuan dalam menentukan pembagian jumlah topik terbaik secara optimal. Semakin tinggi *coherence score* pada sejumlah k topik tertentu, menandakan bahwa untuk setiap topik akan lebih banyak kata yang berkaitan dan topik akan lebih masuk akal (Tang, 2019).

Coherence score dihitung dengan melakukan perbandingan pasangan kata dalam topik tertentu yang akan menghasilkan suatu ukuran kualitas standar dari sebuah topik. Adapun langkah menghitung *coherence score* adalah sebagai berikut:



Gambar 3.3 Ilustrasi Menghitung *Coherence Score*

Sumber: (Röder, Both, & Hinneburg, 2015)

1. *Segmentation* (segmentasi), melakukan segmentasi terhadap sekumpulan kata (t) untuk memperoleh kumpulan *segmented* kata (S),
2. *Probability Calculation*, menghitung nilai probabilitas (P) dari masing-masing kumpulan *segmented* kata (S) berdasarkan frekuensi kata pada korpus.
3. *Confirmation measure*, hasil P dan S akan dihitung pengukuran seberapa bagus kualitas topik berdasarkan suatu metrik tertentu pada setiap partisi untuk memperoleh nilai φ vektor, sehingga didefinisikan sebagai vektor hasil pengukuran yang sudah terkonfirmasi dihasilkan dari confirmation module.

4. *Aggregation* (agregasi), menghitung nilai rata-rata dari semua nilai $\varphi = \varphi_1, \varphi_2, \dots, \varphi_k$ untuk menghasilkan nilai coherence (c). (Röder, Both, & Hinneburg, 2015)

3.12. Word Cloud

Word cloud adalah semacam daftar berbobot untuk representasi visual dari data bahasa atau teks. Visualisasi kata-kata yang lebih sering digunakan dalam teks tampak lebih besar sedangkan kata-kata yang jarang digunakan tampak lebih kecil. *Word cloud* dapat digunakan dalam berbagai aktivitas dan sangat berguna dalam diskusi untuk melihat pratinjau teks dan kosakata (Stamper, 2012). *Word cloud* merupakan cara yang paling populer dan menarik untuk menampilkan data teks dalam bentuk grafik dan dapat menjadi alat yang berguna dalam analisis karena dapat dengan cepat dan mudah menghasilkan penggambaran grafis yang mewakili sebuah teks dengan menyelidiki pola kata atau frasa (DePaolo & Wilkinson, 2014).

BAB IV

METODOLOGI PENELITIAN

4.1. Populasi Penelitian

Populasi yang digunakan dalam penelitian ini adalah seluruh dokumen putusan kasus perceraian yang terdapat di Pengadilan Agama Yogyakarta. Besarnya populasi penelitian sebanyak 6910 dokumen per tanggal 05 April 2022 pada *website* Direktori Putusan Mahkamah Agung Republik Indonesia.

4.2. Tempat dan Waktu Penelitian

Penelitian dilaksanakan pada bulan Desember 2021 sampai April 2022 bertempat di Kota Yogyakarta. Jenis data yang akan digunakan oleh peneliti merupakan data sekunder. Data yang digunakan diperoleh dari Pengadilan Agama Yogyakarta dan *website* Direktori Putusan Mahkamah Agung Republik Indonesia yaitu <https://putusan3.mahkamahagung.go.id/>. Pengambilan data dilakukan pada tanggal 09 - 14 Desember 2021.

4.3. Variabel Penelitian

Variabel yang digunakan dalam penelitian ini adalah alasan pengajuan perceraian, klasifikasi putusan perceraian, dan faktor penyebab perceraian di Pengadilan Agama Yogyakarta tahun 2020.

Tabel 4.1 Variabel Penelitian

No.	Variabel	Definisi Operasional Variabel
1	Alasan pengajuan perceraian	Teks yang memuat alasan / dalil-dalil dari pengadu untuk mengajukan perceraian yang termuat dalam bagian duduk perkara direktori putusan Mahkamah Agung Republik Indonesia.
2	Klasifikasi Putusan Perceraian	Jenis kasus perceraian dalam putusan perceraian.
3	Faktor Penyebab Perceraian	Beberapa faktor yang menyebabkan terjadinya perceraian di Kota Yogyakarta

4.4. Teknik Sampling

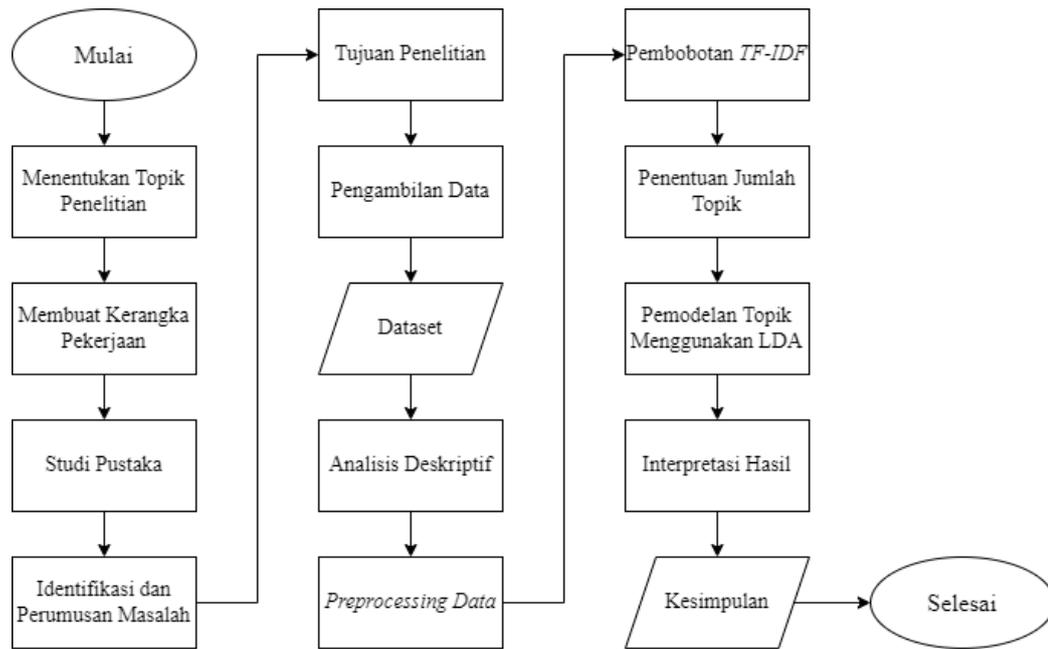
Teknik sampling yang digunakan dalam penelitian ini adalah *Teknik Non Probability Sampling* dengan batasan waktu pada tahun 2020. Sampel yang digunakan adalah dokumen putusan kasus perceraian (alasan pengajuan perceraian yang termuat dalam bagian duduk perkara) di Pengadilan Agama Yogyakarta yang berhasil di putus pada tahun 2020. Banyaknya sampel yang didapat yaitu 685 dokumen.

4.5. Alat dan Cara Organisir Data

Metode analisis data yang digunakan dalam penelitian adalah statistika deskriptif dan pemodelan topik menggunakan metode *Latent Dirichlet Allocation* (LDA). Beberapa *software* yang digunakan untuk membantu analisis dan visualisasi adalah *Microsoft Excel 365* dan *Anaconda3 2021.05 (Python 3.8.8 64-bit)*. Adapun langkah-langkah dan analisa dari penelitian adalah sebagai berikut:

1. Analisis deskriptif digunakan untuk menampilkan representasi atau gambaran umum dari kasus perceraian di Pengadilan Agama Yogyakarta tahun 2020
2. *Preprocessing data* digunakan untuk mengubah teks yang akan dianalisis menjadi lebih sederhana sehingga dapat mewakili dokumen.
3. Pembobotan kata digunakan untuk mengubah data teks menjadi data numerik dengan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF).
4. Visualisasi *topics Coherence* merupakan metode yang digunakan untuk menentukan jumlah topik.
5. Pemodelan topik dengan menggunakan metode *Latent Dirichlet Allocation* (LDA) digunakan untuk mengelompokkan setiap *term* menjadi beberapa topik dari setiap dokumen.

Berikut merupakan langkah-langkah pemodelan topik pada penelitian ini yang digambarkan dalam diagram alir sebagai berikut:



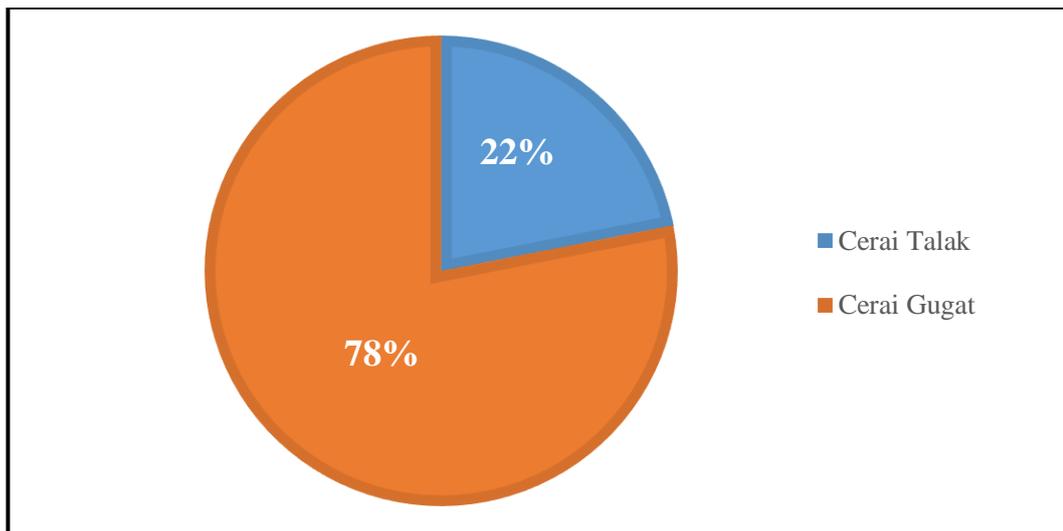
Gambar 4.1 Diagram Alir Penelitian

BAB V

HASIL DAN PEMBAHASAN

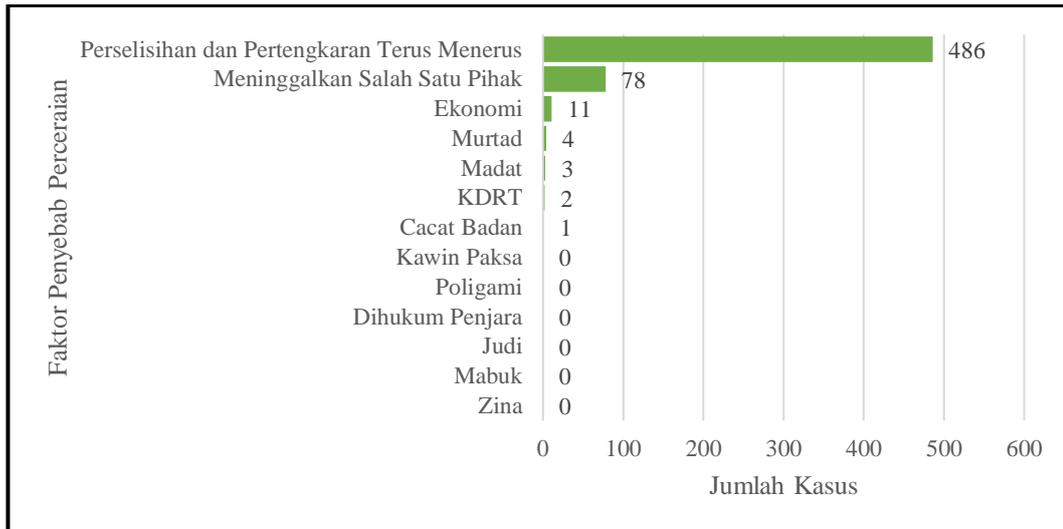
5.1. Analisis Deskriptif

Analisis deskriptif adalah bagian dari ilmu statistika yang hanya mengolah, menyajikan data dengan melihat gambaran secara umum dari data yang didapatkan. Analisis deskriptif memiliki tujuan untuk memberikan gambaran atau mendeskripsikan mengenai suatu data agar tersaji dengan mudah dipahami dan informatif. Analisis deskriptif dalam penulisan ini bertujuan untuk melihat gambaran umum pada data kasus perceraian di Pengadilan Agama Yogyakarta tahun 2020. Data yang digunakan untuk analisis deskriptif merupakan data sekunder yang bersumber dari Pengadilan Agama Yogyakarta.



Gambar 5.1 Persentase Jenis Kasus Perceraian

Gambar 5.1 merupakan diagram lingkaran yang menunjukkan persentase dari perkara perceraian di Pengadilan Agama Yogyakarta dalam kurun waktu satu tahun. Berdasarkan **Gambar 5.1** didapatkan informasi bahwa sebagian besar perceraian di Pengadilan Agama Yogyakarta merupakan cerai gugat yaitu sebesar 78% (sejumlah 503 perkara), sedangkan cerai talak sebesar 22% (sejumlah 142 perkara). Kepala Seksi I Bimbingan Direktorat Jenderal Badan Peradilan Agama, Mahkamah Agung, Hermansyah Hasyim menilai bahwa putusan cerai gugat selalu lebih tinggi dibanding cerai talak yaitu kisaran 60-70% dari jumlah keseluruhan (Hidayat, 2018).



Gambar 5.2 Diagram Faktor Penyebab Perceraian

Gambar 5.2 merupakan gambar diagram batang yang menggambarkan beberapa faktor penyebab perceraian yang berhasil dirangkum oleh Pengadilan Agama Yogyakarta dalam kurun waktu satu tahun. Pada diagram batang tersebut menunjukkan secara runtut faktor yang paling sering menjadi penyebab perceraian. Berdasarkan **Gambar 5.2** didapatkan informasi bahwa faktor perselisihan dan pertengkaran terus-menerus merupakan faktor paling mendominasi yang diusulkan oleh penggugat atau pemohon. Terdapat 486 kasus perceraian yang disebabkan oleh perselisihan dan pertengkaran terus menerus. Kemudian disusul oleh enam faktor lainnya yang menjadi penyebab perceraian di Pengadilan Agama Yogyakarta. Faktor kedua adalah meninggalkan salah satu pihak dengan jumlah 78 kasus, faktor ketiga dengan sebab ekonomi sejumlah 11 kasus. Berdasarkan yurisdiksi Pengadilan Agama seluruh Indonesia faktor penyebab perceraian banyak didominasi oleh faktor perselisihan dan pertengkaran yang terus menerus, ekonomi, dan meninggalkan salah satu pihak (Hidayat, 2018).

5.2. Text Preprocessing

Tahap *text preprocessing* dilakukan untuk membersihkan atau menghilangkan kata-kata dalam dokumen yang tidak dibutuhkan dan mengubah kata menjadi lebih terstruktur. Tahap *text preprocessing* merupakan tahap awal yang dilakukan agar bisa dilanjutkan ke tahap selanjutnya yaitu menganalisis data

teks. Berikut adalah beberapa contoh data yang digunakan dalam penelitian ini, sedangkan sintaks program dan data lengkap dapat dilihat dalam lampiran.

Tabel 5.1 Teks Awal

Teks
<p>Bahwa pada awal perkawinan hubungan kekeluargaan antara Pemohon dengan Termohon masih Harmonis. Namun mulai Tahun 2016 mulai goyah Rumah Tangganya karena sering terjadi perselisihan dan pertengkaran yang terus-menerus yang disebabkan antara lain : Bahwa sejak awal perkawinan, Termohon sering bekerja dari malam sampai pagi sebagai Pemandu Karaoke, sehingga setiap Pemohon menyarankan agar mengganti pekerjaannya yang lain, akan tetapi Termohon tidak memperdulikan saran dan pendapat Pemohon sebagai Suami, sehingga kemudian sering terjadi perselisihan paham dan muncul pertengkaran yang terus-menerus; Bahwa akibat profesi Termohon itu, maka Termohon sering berselingkuh dengan laki-laki lain, hal mana Pemohon sering menemukan foto-foto Termohon dengan laki-laki lain dengan adegan yang tidak layak ditonton, dan Termohon mengaku sebagai selingkuhannya, sehingga setiap hari terjadi pertengkaran dan perselisihan yang terus-menerus; Bahwa berhubung sering bertengkar, maka pada suatu hari dalam Tahun 2017, Termohon mengusir Pemohon dari Rumah tempat tinggal bersama , sehingga mulai Tahun 2017 sudah pisah rumah, dan kemudian menyusul Termohon pulang ke Rumah Orang Tuanya; Bahwa kehidupan Rumah Tangga antara Pemohon dengan Termohon sudah sulit untuk disatukan lagi, sudah 3 tahun berpisah rumah dan sudah tidak ada lagi komunikasi yang baik antara berdua, sehingga telah memenuhi syarat-syarat untuk mengajukan perceraian kepada Pengadilan Agama Yogyakarta dengan Talak Cerai; Bahwa tujuan perkawinan antara Pemohon dengan Termohon yaitu membentuk keluarga yang sakinah dan mawaddah, namun tujuan ini tidak berhasil, sehingga sebaiknya diakhiri dengan perceraian karena yang terjadi sekarang ini adalah penderitaan dan kesengsaraan lahir dan batin;</p>

Tabel 5.1 Merupakan tabel yang berisi contoh dokumen putusan bagian duduk perkara yang mengacu pada sebab pengadu mengajukan kasus perceraian.

Terdapat beberapa tahap yang dilakukan oleh peneliti dalam *text preprocessing* diantaranya *lower casing*, *remove punctuation*, *stopword removal*, dan *tokenizing*.

5.2.1 Lowercasing

Tahap *lowercasing* digunakan untuk mengubah atau menyamaratakan kata yang mempunyai huruf kapital menjadi huruf kecil. Tahap ini bertujuan untuk apabila terdapat kata yang sama namun dalam penulisan huruf berbeda antara huruf kapital dan huruf kecil maka tidak dianggap kata yang berbeda. Contoh hasil dari proses *lowercasing* dapat dilihat pada **Tabel 5.2** berikut.

Tabel 5.2 Hasil Proses *Lowercasing*

Sebelum	Sesudah
<p>Bahwa pada awal perkawinan hubungan kekeluargaan antara Pemohon dengan Termohon masih Harmonis. Namun mulai Tahun 2016 mulai goyah Rumah Tangganya karena sering terjadi perselisihan dan pertengkaran yang terus-menerus yang disebabkan antara lain : Bahwa sejak awal perkawinan, Termohon sering bekerja dari malam sampai pagi sebagai Pemandu Karaoke, sehingga setiap Pemohon menyarankan agar mengganti pekerjaannya yang lain, akan tetapi Termohon tidak memperdulikan saran dan pendapat Pemohon sebagai Suami, sehingga kemudian sering terjadi perselisihan paham dan muncul pertengkaran yang terus-menerus; Bahwa akibat profesi Termohon itu, maka Termohon sering berselingkuh dengan laki-laki lain, hal mana Pemohon sering menemukan</p>	<p>bahwa pada awal perkawinan hubungan kekeluargaan antara pemohon dengan termohon masih harmonis. namun mulai tahun 2016 mulai goyah rumah tangganya karena sering terjadi perselisihan dan pertengkaran yang terus-menerus yang disebabkan antara lain : bahwa sejak awal perkawinan, termohon sering bekerja dari malam sampai pagi sebagai pemandu karaoke, sehingga setiap pemohon menyarankan agar mengganti pekerjaannya yang lain, akan tetapi termohon tidak memedulikan saran dan pendapat pemohon sebagai suami, sehingga kemudian sering terjadi perselisihan paham dan muncul pertengkaran yang terus-menerus; bahwa akibat profesi termohon itu, maka termohon sering berselingkuh dengan laki-laki lain, hal mana pemohon sering menemukan</p>

Sebelum	Sesudah
<p>foto-foto Termohon dengan laki-laki lain dengan adegan yang tidak layak ditonton, dan Termohon mengaku sebagai selingkuhannya, sehingga setiap hari terjadi pertengkaran dan perselisihan yang terus-menerus; Bahwa berhubung sering bertengkar, maka pada suatu hari dalam Tahun 2017, Termohon mengusir Pemohon dari Rumah tempat tinggal bersama, sehingga mulai Tahun 2017 sudah pisah rumah, dan kemudian menyusul Termohon pulang ke Rumah Orang Tuanya; Bahwa kehidupan Rumah Tangga antara Pemohon dengan Termohon sudah sulit untuk disatukan lagi, sudah 3 tahun berpisah rumah dan sudah tidak ada lagi komunikasi yang baik antara berdua, sehingga telah memenuhi syarat-syarat untuk mengajukan perceraian kepada Pengadilan Agama Yogyakarta dengan Talak Cerai; Bahwa tujuan perkawinan antara Pemohon dengan Termohon yaitu membentuk keluarga yang sakinah dan mawaddah, namun tujuan ini tidak berhasil, sehingga sebaiknya diakhiri dengan perceraian karena yang terjadi sekarang ini adalah penderitaan dan kesengsaraan lahir dan batin;</p>	<p>foto-foto termohon dengan laki-laki lain dengan adegan yang tidak layak ditonton, dan termohon mengaku sebagai selingkuhannya, sehingga setiap hari terjadi pertengkaran dan perselisihan yang terus-menerus; bahwa berhubung sering bertengkar, maka pada suatu hari dalam tahun 2017, termohon mengusir pemohon dari rumah tempat tinggal bersama, sehingga mulai tahun 2017 sudah pisah rumah, dan kemudian menyusul termohon pulang ke rumah orang tuanya; bahwa kehidupan rumah tangga antara pemohon dengan termohon sudah sulit untuk disatukan lagi, sudah 3 tahun berpisah rumah dan sudah tidak ada lagi komunikasi yang baik antara berdua, sehingga telah memenuhi syarat-syarat untuk mengajukan perceraian kepada pengadilan agama yogyakarta dengan talak cerai; bahwa tujuan perkawinan antara pemohon dengan termohon yaitu membentuk keluarga yang sakinah dan mawaddah, namun tujuan ini tidak berhasil, sehingga sebaiknya diakhiri dengan perceraian karena yang terjadi sekarang ini adalah penderitaan dan kesengsaraan lahir dan batin;</p>

Berdasarkan **Tabel 5.2** di atas dapat dilihat bahwa perubahan dokumen yang melewati tahap *lowercasing* ditandai dengan huruf yang berwarna hijau yaitu kata yang sebelumnya mempunyai huruf kapital diubah menjadi huruf kecil.

5.2.2 *Remove Punctuation*

Tahap *remove punctuation* digunakan untuk menghapus karakter tanda baca dan angka sehingga dokumen yang dianalisis murni hanya data teks. Karakter tanda baca yang dihapus dapat berupa titik, koma, tanda kurung, dan sebagainya. Contoh dari hasil tahap *remove punctuation* dapat dilihat pada **Tabel 5.3** berikut.

Tabel 5.3 Hasil Proses *Remove Punctuation*

Sebelum	Sesudah
<p>bahwa pada awal perkawinan hubungan kekeluargaan antara pemohon dengan termohon masih harmonis, namun mulai tahun 2016 mulai goyah rumah tangganya karena sering terjadi perselisihan dan pertengkaran yang terus-menerus yang disebabkan antara lain : bahwa sejak awal perkawinan, termohon sering bekerja dari malam sampai pagi sebagai pemandu karaoke, sehingga setiap pemohon menyarankan agar mengganti pekerjaannya yang lain, akan tetapi termohon tidak memperdulikan saran dan pendapat pemohon sebagai suami, sehingga kemudian sering terjadi perselisihan paham dan muncul pertengkaran yang terus-menerus; bahwa akibat profesi termohon itu, maka termohon sering berselingkuh dengan laki-laki lain, hal mana pemohon sering menemukan</p>	<p>bahwa pada awal perkawinan hubungan kekeluargaan antara pemohon dengan termohon masih harmonis namun mulai tahun mulai goyah rumah tangganya karena sering terjadi perselisihan dan pertengkaran yang terus menerus yang disebabkan antara lain bahwa sejak awal perkawinan termohon sering bekerja dari malam sampai pagi sebagai pemandu karaoke sehingga setiap pemohon menyarankan agar mengganti pekerjaannya yang lain akan tetapi termohon tidak memperdulikan saran dan pendapat pemohon sebagai suami sehingga kemudian sering terjadi perselisihan paham dan muncul pertengkaran yang terus menerus bahwa akibat profesi termohon itu maka termohon sering berselingkuh dengan laki laki lain hal mana pemohon sering menemukan</p>

Sebelum	Sesudah
<p>foto-foto termohon dengan laki-laki lain dengan adegan yang tidak layak ditonton, dan termohon mengaku sebagai selingkuhannya, sehingga setiap hari terjadi pertengkaran dan perselisihan yang terus-menerus; bahwa berhubung sering bertengkar, maka pada suatu hari dalam tahun 2017, termohon mengusir pemohon dari rumah tempat tinggal bersama, sehingga mulai tahun 2017 sudah pisah rumah, dan kemudian menyusul termohon pulang ke rumah orang tuanya; bahwa kehidupan rumah tangga antara pemohon dengan termohon sudah sulit untuk disatukan lagi, sudah 3 tahun berpisah rumah dan sudah tidak ada lagi komunikasi yang baik antara berdua, sehingga telah memenuhi syarat-syarat untuk mengajukan perceraian kepada pengadilan agama yogyakarta dengan talak cerai; bahwa tujuan perkawinan antara pemohon dengan termohon yaitu membentuk keluarga yang sakinah dan mawaddah, namun tujuan ini tidak berhasil, sehingga sebaiknya diakhiri dengan perceraian karena yang terjadi sekarang ini adalah penderitaan dan kesengsaraan lahir dan batin;</p>	<p>foto foto termohon dengan laki laki lain dengan adegan yang tidak layak ditonton dan termohon mengaku sebagai selingkuhannya sehingga setiap hari terjadi pertengkaran dan perselisihan yang terus menerus bahwa berhubung sering bertengkar maka pada suatu hari dalam tahun termohon mengusir pemohon dari rumah tempat tinggal bersama sehingga mulai tahun sudah pisah rumah dan kemudian menyusul termohon pulang ke rumah orang tuanya bahwa kehidupan rumah tangga antara pemohon dengan termohon sudah sulit untuk disatukan lagi sudah tahun berpisah rumah dan sudah tidak ada lagi komunikasi yang baik antara berdua sehingga telah memenuhi syarat syarat untuk mengajukan perceraian kepada pengadilan agama yogyakarta dengan talak cerai bahwa tujuan perkawinan antara pemohon dengan termohon yaitu membentuk keluarga yang sakinah dan mawaddah namun tujuan ini tidak berhasil sehingga sebaiknya diakhiri dengan perceraian karena yang terjadi sekarang ini adalah penderitaan dan kesengsaraan lahir dan batin</p>

Berdasarkan **Tabel 5.3** di atas dapat dilihat bahwa perubahan dokumen yang melewati tahap *remove punctuation* ditandai dengan tanda yang berwarna hijau. Pada karakter tanda baca dan angka tersebut kemudian dihilangkan sehingga pada kolom sesudah tidak terdapat karakter-karakter tersebut.

5.2.3 Stopword Removal

Tahap *stopword removal* digunakan untuk menghapus kata-kata yang tidak diperlukan dalam analisis. Beberapa kata penghubung, kata yang sering muncul dalam beberapa dokumen, dan kata yang tidak berkaitan dengan topik yang akan dianalisis. Beberapa kata yang termasuk *stopword removal* dalam analisis teks Bahasa Indonesia adalah “dan”, “dengan”, “atau”, dan lain sebagainya. Contoh dari hasil tahap *stopword removal* dapat dilihat pada **Tabel 5.4** berikut.

Tabel 5.4 Hasil Proses *Stopword Removal*

Sebelum	Sesudah
bahwa pada awal perkawinan hubungan kekeluargaan antara pemohon dengan termohon masih harmonis namun mulai tahun mulai goyah rumah tangganya karena sering terjadi perselisihan dan pertengkaran yang terus menerus yang disebabkan antara lain bahwa sejak awal perkawinan termohon sering bekerja dari malam sampai pagi sebagai pemandu karaoke sehingga setiap pemohon menyarankan agar mengganti pekerjaannya yang lain akan tetapi termohon tidak memperdulikan saran dan pendapat pemohon sebagai suami sehingga kemudian sering terjadi perselisihan paham dan muncul pertengkaran yang terus menerus bahwa akibat profesi termohon itu	perkawinan hubungan kekeluargaan goyah rumah tangganya perselisihan pertengkaran menerus disebabkan perkawinan pemandu karaoke menyarankan mengganti pekerjaannya tidak memedulikan saran pendapat perselisihan paham muncul pertengkaran menerus akibat profesi berselingkuh laki laki menemukan foto foto laki laki adegan tidak layak ditonton mengaku selingkuhannya pertengkaran perselisihan menerus berhubung bertengkar mengusir rumah tinggal pisah rumah menyusul pulang rumah orang tuanya kehidupan rumah tangga sulit disatukan berpisah rumah tidak komunikasi berdua memenuhi syarat syarat mengajukan perceraian talak cerai tujuan perkawinan

Sebelum	Sesudah
<p>maka termohon sering berselingkuh dengan laki laki lain hal mana pemohon sering menemukan foto foto termohon dengan laki laki lain dengan adegan yang tidak layak ditonton dan termohon mengaku sebagai selingkuhannya sehingga setiap hari terjadi pertengkaran dan perselisihan yang terus menerus bahwa berhubung sering bertengkar maka pada suatu hari dalam tahun termohon mengusir pemohon dari rumah tempat tinggal bersama sehingga mulai tahun sudah pisah rumah dan kemudian menyusul termohon pulang ke rumah orang tuanya bahwa kehidupan rumah tangga antara pemohon dengan termohon sudah sulit untuk disatukan lagi sudah tahun berpisah rumah dan sudah tidak ada lagi komunikasi yang baik antara berdua sehingga telah memenuhi syarat syarat untuk mengajukan perceraian kepada pengadilan agama yogyakarta dengan talak cerai bahwa tujuan perkawinan antara pemohon dengan termohon yaitu membentuk keluarga yang sakinah dan mawaddah namun tujuan ini tidak berhasil sehingga sebaiknya diakhiri dengan perceraian karena yang terjadi sekarang ini adalah penderitaan dan kesengsaraan lahir dan</p>	<p>membentuk keluarga tujuan tidak berhasil perceraian penderitaan kesengsaraan</p>

Sebelum	Sesudah
batin	

Berdasarkan **Tabel 5.4** di atas dapat dilihat bahwa perubahan dokumen yang melewati tahap *stopword removal* ditandai dengan kata yang berwarna hijau. Kata-kata yang ditandai kemudian dihilangkan sehingga pada kolom sesudah tidak terdapat kata-kata tersebut.

5.2.4 Lemmatization

Tahap *lemmatization* digunakan untuk menemukan bentuk dasar dari sebuah kata. Misalkan kata “memukul”, “memukuli”, “dipukuli” akan ditransformasi menjadi kata “pukul”. Contoh dari hasil tahap *lemmatization* dapat dilihat pada **Tabel 5.5** berikut.

Tabel 5.5 Hasil Proses *Lemmatization*

Sebelum	Sesudah
perkawinan hubungan kekeluargaan	kawin hubung keluarga goyah rumah
goyah rumah tangganya perselisihan	tangga selisih tengkar kawin pandu
pertengkaran perkawinan pemandu	karaoke saran ganti kerja tidak peduli
karaoke menyarankan mengganti	saran selisih paham muncul tengkar
pekerjaannya tidak memedulikan saran	akibat profesi selingkuh laki laki temu
perselisihan paham muncul	foto foto laki laki adegan tidak layak
pertengkaran akibat profesi	tonton selingkuh tengkar selisih hubung
berselingkuh laki laki menemukan	tengkar usir rumah tinggal pisah rumah
foto foto laki laki adegan tidak layak	susul pulang rumah orang tua hidup
ditonton selingkuhannya pertengkaran	rumah tangga sulit satu pisah rumah
perselisihan berhubung bertengkar	tidak komunikasi dua penuh syarat
mengusir rumah tinggal pisah rumah	syarat aju cerai talak cerai tuju kawin
menyusul pulang rumah orang tuanya	bentuk keluarga tuju tidak hasil cerai
kehidupan rumah tangga sulit	derita sengsara
disatukan berpisah rumah tidak	
komunikasi berdua memenuhi syarat	
syarat mengajukan perceraian talak	
cerai tujuan perkawinan membentuk	
keluarga tujuan tidak berhasil	

Sebelum	Sesudah
perceraian penderitaan kesengsaraan	

Berdasarkan **Tabel 5.5** di atas dapat dilihat bahwa perubahan dokumen yang melewati tahap *lemmatization* ditandai dengan kata yang berwarna hijau. Kata-kata yang ditandai kemudian diubah menjadi kata dasar sehingga pada kolom sesudah tidak terdapat kata-kata yang memiliki imbuhan.

5.2.5 Tokenizing

Tahap *tokenizing* digunakan untuk memisahkan kata pada dokumen menjadi potongan-potongan kata yang akan menjadi entitas sehingga dapat memiliki nilai dalam pembobotan kata atau penyusunan matriks pada tahap selanjutnya. Contoh dari hasil tahap *tokenizing* dapat dilihat pada **Tabel 5.6** berikut.

Tabel 5.6 Hasil Proses *Tokenizing*

Sebelum	Sesudah
kawin hubung keluarga goyah rumah tangga selisih tengkar kawin pandu karaoke saran ganti kerja tidak peduli saran selisih paham muncul tengkar akibat profesi selingkuh laki laki temu foto foto laki laki adegan tidak layak tonton selingkuh tengkar selisih hubung tengkar usir rumah tinggal pisah rumah susul pulang rumah orang tua hidup rumah tangga sulit pisah rumah tidak komunikasi penuh syarat syarat aju cerai talak cerai tuju kawin bentuk keluarga tuju tidak hasil cerai derita sengsara	['kawin', 'hubung', 'keluarga', 'goyah', 'rumah', 'tangga', 'selisih', 'tengkar', 'kawin', 'pandu', 'karaoke', 'saran', 'ganti', 'kerja', 'tidak', 'peduli', 'saran', 'selisih', 'paham', 'muncul', 'tengkar', 'akibat', 'profesi', 'selingkuh', 'laki', 'laki', 'temu', 'foto', 'foto', 'laki', 'laki', 'adegan', 'tidak', 'layak', 'tonton', 'selingkuh', 'tengkar', 'selisih', 'hubung', 'tengkar', 'usir', 'rumah', 'tinggal', 'pisah', 'rumah', 'susul', 'pulang', 'rumah', 'orang', 'tua', 'hidup', 'rumah', 'tangga', 'sulit', 'pisah', 'rumah', 'tidak', 'komunikasi', 'penuh', 'syarat', 'syarat', 'aju', 'cerai', 'talak', 'cerai', 'tuju', 'kawin', 'bentuk', 'keluarga', 'tuju', 'tidak', 'hasil', 'cerai', 'derita', 'sengsara']

Berdasarkan **Tabel 5.6** di atas dapat dilihat bahwa pada kolom sebelum data berbentuk teks yang masih menyatu kemudian pada kolom sesudah susunan kata

Perhitungan pembobotan kata berdasarkan frekuensi kata dalam setiap dokumen. Ketika frekuensi kata lebih sering muncul dalam dokumen tertentu atau kejarangmunculan kata, maka menunjukkan bahwa bobot meningkat dan kata tersebut dianggap penting. Jika kata tersebut sering muncul di banyak dokumen, maka menunjukkan bahwa bobot berkurang yang seberapa umum kata tersebut. Sehingga pembagian setiap dokumen dalam proses analisis penting dilakukan untuk menentukan pembobotan setiap kata sehingga nantinya dapat dihasilkan kata-kata yang relevan.

5.3. Pembobotan kata

Tahap pembobotan kata merupakan tahap setelah *text preprocessing* yang sebelumnya akan dilakukan pengubahan data yang sebelumnya berbentuk teks menjadi angka atau numerik. Dalam analisis *topic modeling*, data yang digunakan harus berskala data numerik. Untuk mengubah data teks yang akan dianalisis menjadi data numerik dapat menggunakan metode pembobotan TF-IDF. Pembobotan TF-IDF merupakan gabungan dari metode *Term Frequency* (TF) dengan metode *Inverse Document Frequency* (IDF). Berikut merupakan beberapa sampel kata yang digunakan untuk menghitung pembobotan kata dan hasil perhitungan nilai TF.

Tabel 5.7 Hasil Perhitungan Nilai TF

Dokumen	abai	abghadhul	abrik	...	cerai	...	zina
1	0	0	0	...	1	...	0
2	0	0	0	...	1	...	0
3	0	0	0	...	3	...	0
4	0	0	0	...	3	...	0
5	0	0	0	...	1	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
684	0	0	0	...	3	...	0
685	0	0	0	...	4	...	0

Berdasarkan perhitungan nilai TF didapatkan matriks berukuran 685×2157 . Dapat dilihat bahwa **Tabel 5.7** yang berisi beberapa contoh hasil yang diperoleh dari nilai TF yang dihitung secara otomatis menggunakan bantuan *software Python*. Nilai TF merupakan frekuensi kemunculan *tern* (t) atau kata pada setiap

dokumen (D). Apabila nilai TF bernilai 0 artinya kemunculan *tern* (t) tidak ada pada dokumen (D), sedangkan apabila bernilai 1 artinya terdapat 1 kemunculan *tern* (t) pada dokumen (D) dan berlaku untuk nilai selanjutnya. Setelah didapatkan nilai TF kemudian dilanjutkan menghitung nilai TF-IDF. Berikut beberapa sampel kata yang digunakan untuk perhitungan nilai TF-IDF.

Tabel 5.8 Hasil Perhitungan Nilai TF-IDF

Dokumen	abai	abghadhul	abrik	...	cerai	...	zina
1	0	0	0	...	1.071081	...	0
2	0	0	0	...	1.071081	...	0
3	0	0	0	...	3.213242	...	0
4	0	0	0	...	3.213242	...	0
5	0	0	0	...	1.071081	...	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
684	0	0	0	...	3.213242	...	0
685	0	0	0	...	4.284322	...	0

Pada **Tabel 5.8** di atas berisi beberapa contoh hasil yang diperoleh dari perhitungan nilai TF-IDF yang dihitung secara otomatis dengan bantuan *software Python*. Nilai TF-IDF merupakan nilai yang didapatkan dari perkalian antara nilai TF dan nilai IDF. Nilai TF-IDF dapat dihitung secara manual dengan menggunakan nilai TF dan nilai IDF. Berikut merupakan sampel perhitungan nilai TF-IDF.

Peneliti menggunakan sampel kata yaitu “cerai” yang akan dijadikan contoh perhitungan nilai TF-IDF. Kata “cerai” mulai muncul pada dokumen ke-1 yang ditandai pada **Tabel 5.7** nilai TF bernilai 1. Kemudian dilanjutkan menghitung nilai IDF atau *inverse* dari DF, nilai IDF dihitung dengan menggunakan nilai DF_t dan nilai DF seperti pada **Persamaan 3.3**. Nilai IDF dipengaruhi oleh nilai DF_t , semakin rendah nilai DF_t maka semakin tinggi nilai IDF dan sebaliknya. Nilai DF_t merupakan nilai dari banyaknya dokumen yang mengandung *tern* (t) atau yang mengandung kata “cerai” yaitu sebanyak 638 dokumen. Sedangkan nilai DF merupakan nilai dari banyaknya dokumen yang digunakan dalam penelitian yaitu sebanyak 685 dokumen. Berikut contoh perhitungan nilai IDF.

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

$$IDF_{cerai} = \ln\left(\frac{685}{638}\right) = 0.07108055492$$

Setelah didapatkan nilai TF, DF, dan IDF kemudian dilanjutkan dengan menghitung nilai FT-IDF yang akan digunakan untuk pembobotan kata. Nilai TF-IDF didapatkan dengan mengalikan nilai TF dan nilai IDF atau berdasar pada **Persamaan 3.6**. Berikut merupakan contoh perhitungan nilai TF-IDF pada kata “cerai” dengan nilai TF sebesar 1.

$$w = TF * (IDF + 1)$$

$$w = 1 * (0.07108055492 + 1)$$

$$w = 1 * (1.07108055492)$$

$$w = 1.07108055492$$

$$w \approx 1.071081$$

Berdasarkan hasil perhitungan pembobotan kata di atas dapat disimpulkan bahwa kata “cerai” pada dokumen ke 1 mempunyai pembobotan kata TF-IDF sebesar 1,071081. Hasil perhitungan manual dari nilai TF-IDF dari kata “cerai” dapat dirangkum pada tabel sebagai berikut.

Tabel 5.9 Hasil Pembobotan Kata Cerai

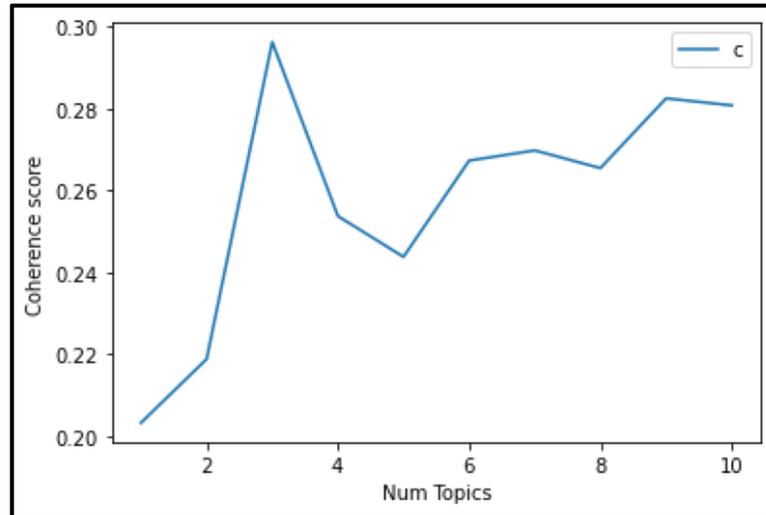
<i>Term (t)</i>	<i>TF</i>	<i>DF</i>	<i>DF_t</i>	<i>IDF</i>	<i>TF – IDF</i>
	<i>D₁</i>				<i>D₁</i>
cerai	1	685	638	0.071081	1.071081

Setelah dilakukan perhitungan pembobotan kata secara manual, kemudian dapat dibandingkan dengan perhitungan secara otomatis dengan menggunakan bantuan *software Python*. Hasil perhitungan dapat dilihat pada lampiran, dapat dilihat bahwa hasil perhitungan secara manual dan otomatis menghasilkan nilai yang sama.

5.4. Topic Modeling dengan Latent Dirichlet Allocation (LDA)

Dalam menentukan jumlah topik dalam *topic modeling* dapat dilakukan dengan melihat visualisasi pada gambar grafik *topics coherence score* (nilai *coherence*). Nilai *coherence* merupakan suatu ukuran nilai yang dapat digunakan untuk mengevaluasi *topic modeling*, dimana dengan melihat nilai *coherence* yang

semakin tinggi maka model yang dihasilkan tersebut semakin baik. Sehingga jumlah topik dengan nilai *coherence* tertinggi yang akan digunakan dalam analisis *topic modeling*. Grafik yang dihasilkan dari nilai *coherence* tergolong fluktuatif, berikut grafik yang dihasilkan dimana topik dimulai dari 0-11.



Gambar 5.4 Grafik Nilai *Coherence*

Berdasarkan **Gambar 5.4** didapatkan informasi bahwa grafik nilai *coherence* memiliki pola yang tidak stabil atau fluktuatif. Penggunaan nilai limit untuk menentukan jumlah topik tergantung pada peneliti. Peneliti memutuskan untuk menggunakan grafik nilai *coherence* dengan limit topik 0-11 dengan nilai *coherence* tertinggi pada jumlah topik sebanyak 3. Beberapa grafik tersebut hanya menggambarkan nilai *coherence* saja, untuk memudahkan melihat nilai pada grafik nilai *coherence* dengan limit 0-11 dapat dilihat pada **Tabel 5.10** berikut.

Tabel 5.10 Hasil Nilai *Coherence*

Num Topics	Coherence Value	Num Topics	Coherence Value
1	0.20322	6	0.26727
2	0.21878	7	0.26976
3	0.29627	8	0.26544
4	0.25372	9	0.28248
5	0.24376	10	0.28077

Hasil dari nilai *coherence* pada **Tabel 5.10** menunjukkan bahwa nilai *coherence* paling tinggi menggunakan limit 0-11 dengan nilai sebesar 0.29627 dengan jumlah topik sebanyak 3. Jumlah topik yang telah ditentukan berdasarkan

nilai *coherence* dengan nilai terbaik tersebut akan menjadi acuan untuk membuat model topik selanjutnya, sehingga dalam penelitian akan digunakan sebanyak tiga model topik.

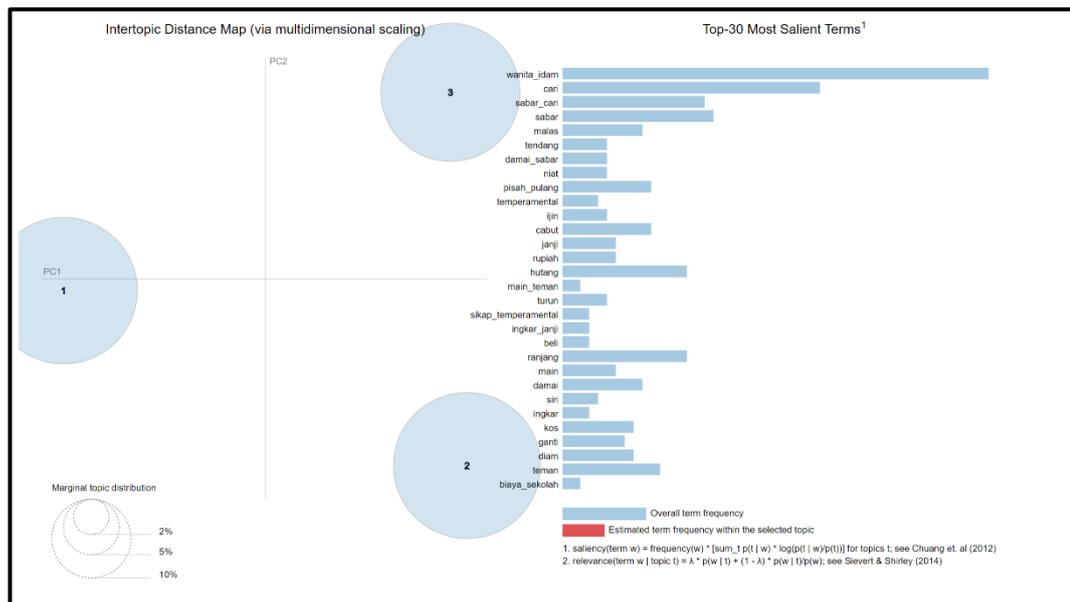
Setelah didapatkan jumlah topik berdasarkan nilai *coherence*, kemudian akan didapatkan model LDA berdasarkan jumlah topik yang digunakan yaitu 3. Analisis *topic modeling* tersebut dapat divisualisasikan dengan sebutan *pyLDAvis*. *pyLDAvis* merupakan suatu visualisasi dari metode LDA yang akan menggambarkan keterkaitan antar kata yang dihasilkan dan kemudian dapat disimpulkan menjadi suatu topik tertentu. Visualisasi dengan *pyLDAvis* didominasi menjadi dua bagian yaitu visualisasi dari letak topik dan visualisasi 30 terminologi yang paling relevan. Letak topik secara keseluruhan dibuat berdasarkan pemetaan jarak antar topik (*intertopic distance map*) menggunakan *multidimensional scaling*. Ukuran suatu kelompok topik dapat menjelaskan seberapa penting topik tersebut terhadap kata, apabila suatu topik divisualisasikan berukuran besar maka topik tersebut lebih relevan terhadap kata dan sebaliknya. Untuk memvisualisasikan letak topik dibagi menjadi 2 bagian sumbu *Principal Component* (PC) yaitu PC1 dan PC2 sehingga terbentuk kuadran-kuadran dengan lingkaran topik yang berdekatan artinya memiliki hubungan kuat. Pada lingkaran yang saling berdekatan tersebut terdapat beberapa kata yang cenderung sama. Model topik yang baik memiliki artian kata-kata yang berbeda antara topik satu dengan lainnya. Berikut adalah nilai PC untuk setiap topik yang digunakan.

Tabel 5.11 Hasil Nilai PC

Topik	PC1	PC2
1	-0.03377	-0.00101
2	0.017948	-0.02359
3	0.015825	0.024602

Pada **Tabel 5.11** di atas didapatkan nilai PC untuk ketiga topik yang divisualisasi menggunakan *pyLDAvis*. Nilai PC1 digunakan sebagai titik koordinat pada sumbu X, sedangkan nilai PC2 digunakan sebagai titik koordinat pada sumbu Y. Dengan melihat hasil nilai PC1 dan PC2 didapatkan informasi bahwa letak topik 1 berada di kuadran III yang ditandai dengan nilai pada sumbu X negatif dan nilai pada sumbu Y negatif, letak topik 2 berada di kuadran IV, dan

letak topik 3 berada di kuadran I. Berikut merupakan hasil visualisasi dari *pyLDavis* berdasarkan tiga topik yang diperoleh.



Gambar 5.5 Visualisasi Ketiga Topik

Gambar 5.5 merupakan gambar visualisasi dari ketiga topik yang didapatkan. Pada bagian kanan memperlihatkan letak antar topik yang menyebar di berbagai kuadran yang berbeda, sehingga dapat dikatakan hubungan antar topik tergolong lemah karena jarak topik satu dengan yang lain jauh. Pada bagian kanan **Gambar 5.5** memperlihatkan *bar chart top 30 global term* paling relevan pada suatu topik tertentu. Diagram batang yang berwarna biru diartikan sebagai keseluruhan *term frequency* dalam *corpus*, sedangkan diagram batang yang berwarna merah diartikan sebagai estimasi *term frequency* terhadap suatu topik tertentu.

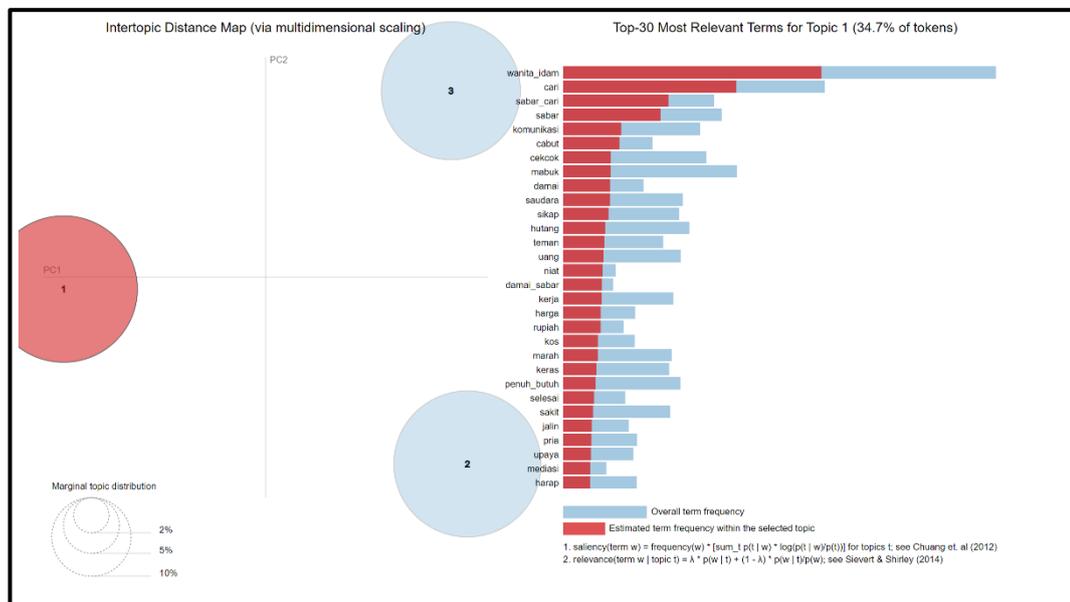


Gambar 5.6 Word Cloud Ketiga Topik

Gambar 5.6 menunjukkan *word cloud* 30 *term* yang paling relevan berdasarkan 10 *term* yang paling relevan pada masing-masing topik. *Word cloud* menunjukkan visualisasi *term* yang mempunyai nilai probabilitas tinggi dalam model. Apabila nilai probabilitas suatu *term* lebih besar dibandingkan dengan *term* lainnya maka *term* tersebut divisualisasikan lebih menonjol daripada yang lain pada *word cloud*. Misalkan *term* “wanita idam” divisualisasikan paling menonjol sehingga didapatkan bahwa kata tersebut memiliki nilai probabilitas yang paling tinggi.

5.4.1 Model LDA Topik ke-1

Pada topik ke-1 didapatkan visualisasi dengan menggunakan *pyLDAvis* sebagai berikut:



Gambar 5.7 Visualisasi topik ke-1

Berdasarkan **Gambar 5.7** dapat dilihat bahwa topik 1 divisualisasikan dengan lingkaran yang berwarna merah. Nilai *marginal topic distribution* pada topik 1 sebesar 34,7% yang diartikan sebagai tingkat pentingnya topik 1 terhadap *corpus*. Pada sisi kanan terdapat diagram batang yang divisualisasikan berwarna merah dan biru pada 30 terminologi. Pada diagram tersebut secara runtut mengurutkan topik yang paling relevan, dengan *term* yang paling relevan yaitu “wanita idam”, “cari”, “sabar cari” kemudian disusul dengan *term* lainnya. Kemudian model pada topik 1 dapat dilihat pada **Tabel 5.12** berikut.

Tabel 5.12 Model Topik 1

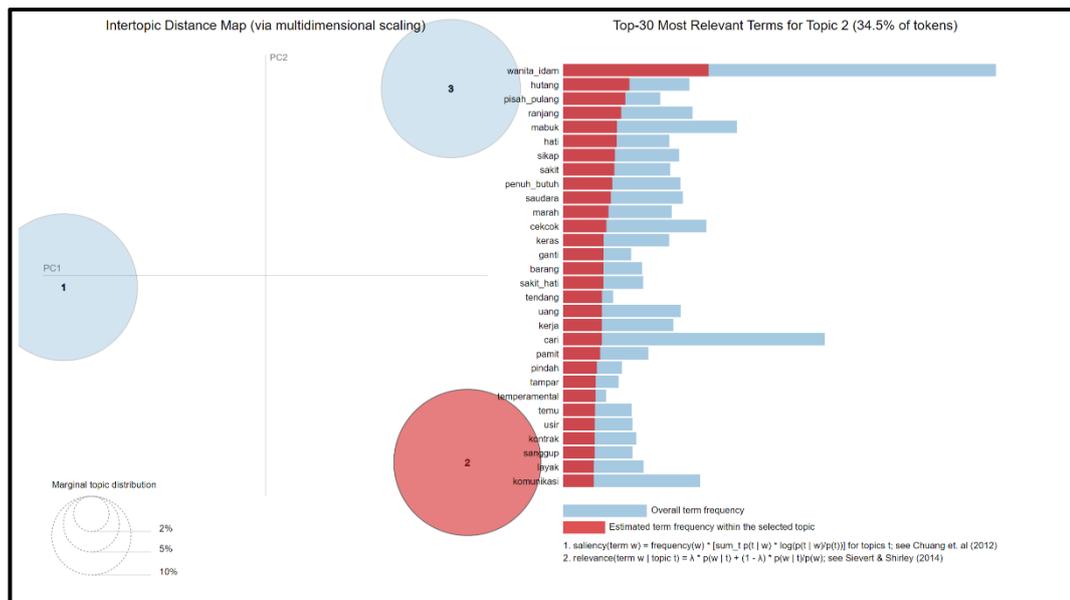
Model Topik 1
Topic: Word: 0.038*"wanita_idam" + 0.026*"cari" + 0.016*"sabar_cari" + 0.014*"sabar" + 0.009*"komunikasi" + 0.008*"cabut" + 0.007*"cekcek" + 0.007*"mabuk" + 0.007*"damai" + 0.007*"saudara"

Pada **Tabel 5.12** yang merupakan hasil dari model pada topik 1, didapatkan 10 *term* teratas yang memiliki peluang penyebutan paling besar. Diperoleh *term* yang mempunyai probabilitas paling besar pada topik 1 adalah “wanita idam” dengan probabilitas sebesar 0.038, diikuti dengan *term* “cari” dengan probabilitas sebesar 0.026, dan *term* “sabar cari” dengan probabilitas 0.016.

Berdasarkan model LDA yang diperoleh dan *pyLDAvis* pada topik 1 maka peneliti menyimpulkan bahwa topik bahasan yang sering muncul dan berkaitan satu dengan yang lain adalah mengenai faktor terjadinya perceraian yaitu meninggalkan salah satu pihak. Simpulan tersebut didasari pada *term* yang relevan yaitu “Wanita idam”, “cari”, “sabar cari”, “sabar”, “damai”, “damai sabar”, dan “pria”.

5.4.2 Model LDA Topik ke-2

Pada topik ke-2 didapatkan visualisasi dengan menggunakan *pyLDAvis* sebagai berikut:



Gambar 5.8 Visualisasi topik ke-2

Berdasarkan **Gambar 5.8** dapat dilihat bahwa topik 2 divisualisasikan dengan lingkaran yang berwarna merah. Nilai *marginal topic distribution* pada topik 2 sebesar 34,5% yang diartikan sebagai tingkat pentingnya topik 2 terhadap *corpus*. Pada sisi kanan terdapat diagram batang yang divisualisasikan berwarna merah dan biru pada 30 terminologi. Pada diagram tersebut secara runtut mengurutkan topik yang paling relevan, dengan *term* yang paling relevan yaitu “wanita idam”, “hutang”, dan “pisah pulang” kemudian disusul dengan *term* lainnya. Kemudian model pada topik 2 dapat dilihat pada **Tabel 5.13** berikut.

Tabel 5.13 Model Topik 2

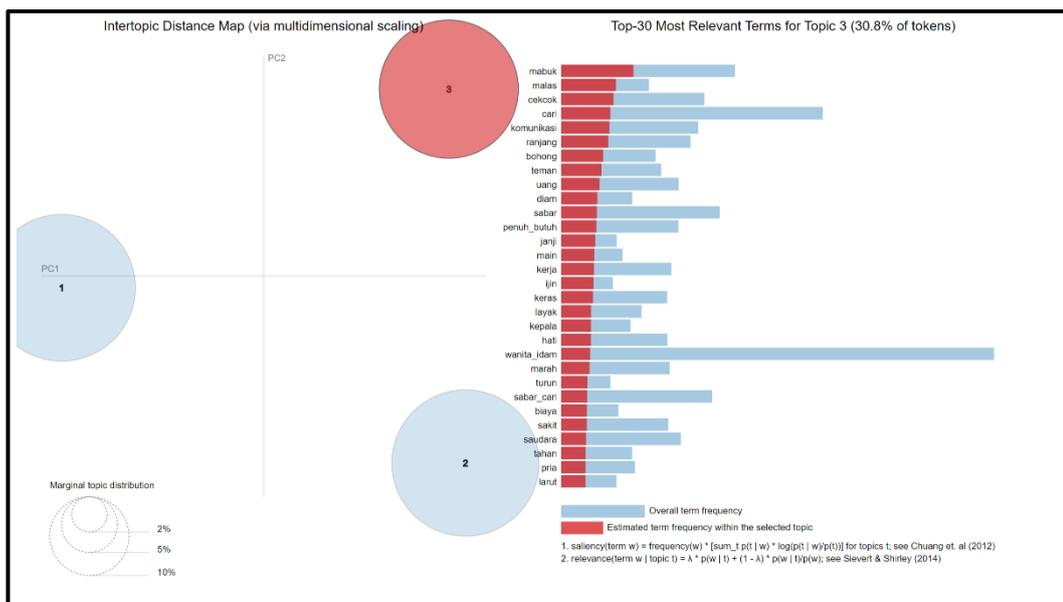
Model Topik 2
Topic: Word: 0.022*"wanita_idam" + 0.010*"hutang" + 0.009*"pisah_pulang" + 0.009*"ranjang" + 0.008*"mabuk" + 0.008*"hati" + 0.008*"sikap" + 0.008*"sakit" + 0.007*"penuh_butuh" + 0.007*"saudara"

Pada **Tabel 5.13** yang merupakan hasil dari model pada topik 2, didapatkan 10 *term* teratas yang memiliki peluang penyebutan paling besar. Diperoleh *term* yang mempunyai probabilitas paling besar pada topik 2 adalah “wanita idam” dengan probabilitas sebesar 0.022, diikuti dengan *term* “hutang” dengan probabilitas sebesar 0.010 dan *term* “pisah pulang” dengan probabilitas 0.009.

Berdasarkan model LDA yang diperoleh dan *pyLDAvis* pada topik 2 maka peneliti menyimpulkan bahwa topik bahasan yang sering muncul dan berkaitan satu dengan yang lain adalah mengenai alasan perceraian yaitu masalah ekonomi. Simpulan tersebut didasari pada *term* yang relevan yaitu “hutang”, “penih butuh”, “uang”, dan “kerja”.

5.4.3 Model LDA Topik ke-3

Pada topik ke-3 didapatkan visualisasi dengan menggunakan *pyLDAvis* sebagai berikut:



Gambar 5.9 Visualisasi topik ke-3

Berdasarkan **Gambar 5.9** dapat dilihat bahwa topik 3 divisualisasikan dengan lingkaran yang berwarna merah. Nilai *marginal topic distribution* pada topik 3 sebesar 30.8% yang diartikan sebagai tingkat pentingnya topik 3 terhadap *corpus*. Pada sisi kanan terdapat diagram batang yang divisualisasikan berwarna merah dan biru pada 30 *term*. Pada diagram tersebut secara runtut mengurutkan topik yang paling relevan, dengan *term* yang paling relevan yaitu “mabuk”, “malas”, kemudian disusul dengan *term* lainnya. Kemudian model pada topik 3 dapat dilihat pada **Tabel 5.14** berikut.

Tabel 5.14 Model Topik 3

Model Topik 3
Topic: Word: 0.012*"mabuk" + 0.009*"malas" + 0.009*"cekcok" + 0.008*"cari" + 0.008*"komunikasi" + 0.008*"ranjang" + 0.007*"bohong" + 0.007*"teman" + 0.006*"uang" + 0.006*"diam"

Pada **Tabel 3.14** yang merupakan hasil dari model pada topik 3, didapatkan 10 *term* teratas yang memiliki peluang penyebutan paling besar. Diperoleh *term* yang mempunyai probabilitas paling besar pada topik 3 adalah “mabuk” dengan probabilitas sebesar 0.012, *term* “malas” dengan probabilitas sebesar 0.009 dan diikuti dengan *term* “cekcok” dengan probabilitas sebesar 0.009.

Berdasarkan model LDA yang diperoleh dan *pyLDAvis* pada topik 3 maka peneliti menyimpulkan bahwa topik bahasan yang sering muncul dan berkaitan

satu dengan yang lain adalah mengenai alasan perceraian yaitu perlakuan salah satu pihak. Simpulan tersebut didasari pada *term* yang relevan yaitu “malas”, “main”, “mabuk”, “cekcok”, “bohong”, “marah”, dan “diam”.

Berdasarkan hasil penelitian didapatkan hasil beberapa topik berkaitan tentang faktor penyebab perceraian berhubungan dengan data yang telah ada. Pada data dari Pengadilan Agama Yogyakarta faktor yang paling menonjol dari penyebab perceraian adalah perselisihan pertengkaran, meninggalkan salah satu pihak dan ekonomi. Hal tersebut dapat digolongkan selaras dengan hasil penelitian yaitu pada topik 1 selaras dengan faktor perceraian kedua dari informasi Pengadilan Agama Yogyakarta yaitu meninggalkan salah satu pihak, topik 2 selaras dengan faktor perceraian ketiga yaitu masalah ekonomi, dan topik 3 merupakan topik tambahan yang ditemukan dari pemodelan topik yaitu perlakuan salah satu pihak.

Analisis pemodelan topik memiliki kelebihan dibandingkan informasi dari Pengadilan Agama Yogyakarta yaitu peneliti dapat menemukan topik tersembunyi dari faktor perceraian yang tidak tercantum dalam analisis deskriptif. Topik tersebut misalnya mengenai faktor perlakuan salah satu pihak merupakan faktor yang tidak terdapat pada informasi dari Pengadilan Agama Yogyakarta, hasil tersebut dapat menjadi informasi baru yang lebih spesifik. Selain itu pada informasi Pengadilan Agama Yogyakarta sebagian besar faktor perceraian mengenai perselisihan pertengkaran yang memungkinkan faktor tersebut terlalu umum untuk menjadi salah satu faktor perceraian.

BAB VI

PENUTUP

5.5. Kesimpulan

Setelah peneliti melakukan penelitian dan mendapatkan hasil, maka dapat diambil kesimpulan sebagai berikut:

1. Gambaran umum pada data putusan perceraian di Pengadilan Agama Yogyakarta pada tahun 2020 yaitu didapatkan informasi bahwa sebagian besar merupakan perkara cerai gugat dengan persentase 78%. Kemudian faktor penyebab perceraian yang paling mendominasi adalah perselisihan dan pertengkarannya terus-menerus terdapat 486 perkara dan disusul oleh enam faktor lainnya yaitu meninggalkan pihak lain, ekonomi, murtad, madat, KDRT, dan cacat badan.
2. Hasil implementasi *Topic Modeling* mengenai faktor perceraian dengan metode *Latent Dirichlet Allocation* (LDA) berdasarkan alasan pengajuan perceraian pada dokumen putusan perceraian di Pengadilan Agama Yogyakarta tahun 2020 diperoleh jumlah topik sebanyak 3 dengan nilai *coherence* 0.29627. Topik yang dihasilkan dari pemodelan topik yaitu sebagai berikut:
 - Topik 1 cenderung membahas mengenai meninggalkan salah satu pihak.
 - Topik 2 cenderung membahas mengenai masalah ekonomi.
 - Topik 3 cenderung membahas mengenai perlakuan salah satu pihak.

5.6. Saran

Berdasarkan hasil penelitian, maka penulis memberikan beberapa saran yaitu sebagai berikut:

1. Data penelitian ini hanya menggunakan dokumen putusan pada tahun 2020, dan hanya dikhususkan merujuk ke faktor penyebab perceraian. Diharapkan pada penelitian selanjutnya dapat menambah data dan selang waktu yang lebih lama sehingga hasil yang didapatkan lebih baik.
2. Dapat melakukan penyempurnaan pada tahap *text preprocessing* yaitu pada bagian *stopwords* dengan menyesuaikan kata yang dapat dihilangkan sehingga tidak menghilangkan makna kata yang terkandung.

3. Pada penelitian ini peneliti hanya menggunakan metode *Latent Dirichlet Allocation* (LDA). Untuk penelitian selanjutnya diharapkan dapat mengembangkan analisis dengan menggunakan atau menambah metode lainnya.

DAFTAR PUSTAKA

- Ali, Z. (2006). *Hukum Perdata Islam di Indonesia*. Jakarta: Sinar Grafika.
- 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.
- Annisa, R., Surjandari, I., & Zulkarnain. (2019). Opinion Mining on Mandalika Hotel Reviews Using Latent Dirichlet Allocation . *Elsevier*.
- Assidiq, Y. (2019, Januari 18). *Kenaikan Angka Perceraian di Yogyakarta tidak Signifikan*. Diambil kembali dari republica.co.id: <https://www.republika.co.id/>
- Blei, D. M. (2012). Probabilistic Topic Models. *Communication of the ACM*.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research* 3, 993-1022.
- BPS. (2017). *Indeks Kebahagiaan 2017*. Jakarta: Badan Pusat Statistik .
- Brett, M. R. (2012, December 12). *Journal of Digital Humanities*. Diambil kembali dari Topic Modeling: A Basic Introduction: <http://journalofdigitalhumanities.org/>
- Campbell, J. C., Hindle, A., & Stroulia, E. (2014). Latent Dirichlet Allocation: Extracting Topics from Software Engineering Data.
- Cendana, M., & Permana, S. D. (2019). PRA-PEMROSESAN TEKS PADA GRUP WHATSAPP UNTUK PEMODELAN TOPIK. *Jurnal Mantik Penusa*, 112.
- DePaolo, C. A., & Wilkinson, K. (2014). Get Your Head into the Clouds: Using Word Clouds for Analyzing Qualitative Assessment Data. *TechTrends*.
- ER, N. A. (2021). Implementasi Latent Dirichlet Allocation (LDA) untuk Klasterisasi Cerita Berbahasa Bali. *Jurnal Teknologi Informasi dan Ilmu Komputer*.
- Feldman, Ronen, & Sanger, J. (2007). *The Text Mining Handbook*. Cambridge: Cambridge University Press.
- Hadi, A. F., W., D. B., & Hasan, M. (2017). Text Mining pada Media Sosial Twitter (Studi Kasus: Masa Tenang Pilkada DKI 2017 Putaran 2). *Seminar Nasional Matematika dan Aplikasinya*.

- Hermawan, I. (2018). Analisis Putusan Pengadilan Agama Karawang Tentang Cerai Gugat Karena Pelanggaran Taklik Talak (Studi Perkara No. 0554/Pdt.G.2015/ PA.Krw). *Jurnal Al-Afkar 1*, 182-195.
- Hidayat, R. (2018, Juni 18). *Melihat Tren Perceraian dan Dominasi Penyebabnya*. Diambil kembali dari Hukumonline.com: <https://www.hukumonline.com/>
- Hikmah, N. (2018). Pemanfaatan Text Mining dalam Pencarian Ayat Al Quran menggunakan TF-IDF dan Cosine Similarity. *Antartika*, 13-22.
- Hudaya, C. S., Fakhrurroja, H., & Alamsyah, A. (2019). ANALISIS PERSEPSI KONSUMEN TERHADAP BRAND GO-JEK PADA MEDIA SOSIAL TWITTER MENGGUNAKAN METODE SENTIMENT ANALYSIS DAN TOPIC MODELLING. *Jurnal Mitra Manajemen (JMM Online)*, 667.
- Ingason, A. K., Helgadottir, S., Loftsson, H., & Rognvaldsson, E. (2008). A Mixed Method Lemmatization Algorithm Using a Hierarchy of Linguistic Identities (HOLI). *Aarne Ranta (Eds.), Advances in Natural Language Processing*.
- Kabiru, I. N., & Sari, P. K. (2019). Analisa Konten Media Sosial E-Commerce Pada Instagram Menggunakan Metode Sentimen Analysis Dan LDA-Based Topic Modeling (Studi Kasus : Shopee Indonesia). *e-Proceeding of Management*, 14.
- Karmayasa, O., & Mahendra, I. B. (2012). IMPLEMENTASI VECTOR SPACE MODEL DAN BEBERAPA NOTASI METODE TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF) PADA SISTEM TEMU KEMBALI INFORMASI. *Jurusan Ilmu Komputer, Universitas Udayana*.
- Ks, M. (2009). *Romantika Perkawinan di Indonesia*. Yogyakarta: DPPAI-UII.
- Latief, D. (1985). *Aneka Hukum Perceraian di Indonesia*. Jakarta: Ghalia Indonesia.
- Lestari. (2019). INISIASI NATURAL LANGUAGE PROCESSING (NLP) DAN KLASIFIKASI JENIS WISATA KULINER UNTUK PROGRAM

- CHATBOT, (Studi Kasus : Informasi Wisata Kuliner Daerah Istimewa Yogyakarta). Yogyakarta: Statistika UII.
- Merchant, K., & Pande, Y. (2018). NLP Based Latent Semantic Analysis for Legal Text Summarization. *IEEE*.
- Nirenburg, S. (2009). *Language Engineering for Lesser-Studied Languages*. Amsterdam: IOS Press.
- Nurjannah, M., Hamdani, & Astuti, I. F. (2013). PENERAPAN ALGORITMA TERM FREQUENCY-INVERSE DOCUMENT FREQUENCY (TF-IDF) UNTUK TEXT MINING. *Jurnal Informatika Mulawarman*.
- Pathana, A. F., & Prakash, C. (2021). Unsupervised Aspect Extraction Algorithm for opinion mining using topic modeling. *Global Transitions Proceedings*.
- Putra, I. M., & Kusumawardani, R. P. (2017). Analisis Topik Informasi Publik Media Sosial di Surabaya Menggunakan Pemodelan Latent Dirichlet Allocation (LDA). *Jurnal Teknik ITS*.
- Rashif, F., Nirvana, G. I., Noor, M. A., & Rakhmawati, N. A. (2021). Implementasi LDA untuk Pengelompokan Topik Cuitan Akun Bot Twitter bertagar #Covid-19. *Cogito Smart Journal*.
- Röder, M., Both, A., & Hinneburg, A. (2015). Exploring the Space of Topic Coherence Measures. *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, 399–408.
- Sanjaya, S., & Absar, E. A. (2015). Pengelompokan Dokumen Menggunakan Winnowing Fingerprint dengan Metode K-Nearest Neighbour. *Jurnal CoreIT*, 52.
- Setijohatmo, U. T., Rachmat, S., Susilawati, T., & Rahman, Y. (2020). Analisis Metoda Latent Dirichlet Allocation untuk Klasifikasi Dokumen Laporan Tugas Akhir Berdasarkan Pemodelan Topik. *Prosiding The 11th Industrial Research Workshop and National Seminar*.
- Sindar, A., & Sitorus, M. (2020). Machine Learning Prediksi Karakter Pengguna Hashtag (#) Bahasa Generasi Milenial di Sosial Media. *Indonesia Journal of Applied Informatics*, 167-171.

- 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*, 171-186.
- Socrates, I. G., Akbar, A. L., & Akbar, M. S. (2016). Optimasi Naive Bayes Dengan Pemilihan Fitur Dan Pembobotan Gain Ratio. *Lontar Komputer : Jurnal Ilmiah Teknologi Informasi*, 23-24.
- Soimin, S. (2016). *Kitab Undang-Undang Hukum Perdata*. Jakarta: Sinar Grafika.
- Stamper, S. (2012). Word Clouds. *TESOL International Association*.
- Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012). Exploring Topic Coherence over many models and many topics. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 952-961.
- Sudarsono. (2005). *Hukum Perkawinan Nasional*. Jakarta: Rineka Cipta.
- Sugiyono. (2009). *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta.
- Tanaka, M., & Okutomi, M. (2014). A novel inference of a restricted boltzmann machine. 22nd International Conference on Pattern Recognition (ICPR). *Stockholm: IEEE Computer Society.*, 1526-1531.
- Tang, F. (2019, July 14). *Beginner's Guide to LDA Topic Modeling with R*. Diambil kembali dari Toward Data Science: <https://towardsdatascience.com/>
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2011). *Probability & Statistics for Engineers & Scientists 9th Ed*. USA: Pearson.
- Wijaya, E. (2013). Analisis Penggunaan Algoritma Breadth First Search Dalam Konsep Artificial Intelligence. *Wijaya, E*, 18-26.
- Xue, J., Chen, J., Chen, C., Zheng, C., Li, S., & Zhu, T. (2020). Public discourse and sentiment during the COVID 19 pandemic: Using Latent Dirichlet Allocation for topic modeling on Twitter. *PLOS ONE*.
- Yoren. (2018). Perbandingan Raw TF dan Binary TF pada Sistem Pencarian di Situs Museum Wayang Kekayon Yogyakarta. *Skripsi*.

Perundang-Undangan

Undang-Undang Nomor 1 Pasal 1 Tahun 1974

Undang-Undang Nomor 1 Pasal 39 ayat 1 dan 2 Tahun 1974

Undang-Undang Nomor 50 Pasal 54 Tahun 2009

LAMPIRAN

Lampiran 1	Surat Izin Penelitian	https://bit.ly/SuratElen
Lampiran 2	Data Perkara Perceraian Pengadilan Agama Yogyakarta Tahun 2020	https://bit.ly/DataPerkaraElen
Lampiran 3	Data Teks Putusan Perceraian	https://bit.ly/DataTeksElen
Lampiran 4	<i>Script dan Output Topic Modeling</i>	https://bit.ly/ScriptElen