

BAB I

PENDAHULUAN

A. Latar Belakang

Perguruan tinggi sebagai salah satu wadah pencetak generasi muda yang berpendidikan dimana harus mampu mempersiapkan lulusan mahasiswa yang siap pakai dan siap guna bagi diri dan masyarakat nantinya. Parameter pencapaian keberhasilannya sebagai anggota masyarakat adalah memiliki kemampuan akademis dan profesional yang dapat menerapkan ilmu yang telah dipelajari selama mengikuti perkuliahan. Salah satunya adalah siap dalam menghadapi tantangan global dan persaingan dalam dunia Usaha dan dunia Industri.

Kerja praktek atau magang merupakan salah satu kegiatan yang dilakukan oleh mahasiswa untuk mendapatkan pengalaman kecerdasan intelektual sebelum memasuki dunia kerja. Pada proses magang tersebut mahasiswa mendapatkan peluang untuk melakukan riset atau penelitian yang nantinya dapat dijadikan bahan referensi skripsi. Kegiatan magang juga dilakukan pada kampus Universitas Ahmad Dahlan Yogyakarta, selain sebagai kegiatan implementasi ilmu selama diperkuliahan, kegiatan magang juga menjadi mata kuliah yang termasuk dalam kartu rencana studi, khususnya pada mahasiswa semester tujuh (7).

Teknologi Informasi dan Komunikasi (TIK) saat ini sangat berkembang di masyarakat, secara umum Teknologi Informasi adalah sebuah teknologi yang digunakan untuk mengolah data, meliputi didalamnya, Memproses, mendapatkan, menyusun, menyimpan dan memanipulasi data dengan berbagai macam cara dan prosedur guna menghasilkan informasi yang berkualitas dan bernilai guna tinggi. Perkembangan Teknologi Informasi dan Komunikasi seakan telah mendarah daging di dalam setiap manusia di era sekarang ini, termasuk dalam dunia pendidikan, menuntut dunia pendidikan untuk selalu dan senantiasa menyesuaikan perkembangan teknologi terhadap usaha dalam

peningkatan mutu pendidikan, terutama penyesuaian penggunaan teknologi informasi dan komunikasi bagi pendidikan khususnya dalam proses pembelajaran, dimana saat ini sedang tren dengan penggunaan e-(elektronik) dalam setiap proses pembelajaran, seperti e-education, e-learning, e- government serta aplikasi-aplikasi tutor pembelajaran dan yang lain sebagainya. Pemerintah Kota Yogyakarta telah mengintegrasikan informasi publik wilayah kota ke dalam aplikasi Jogja Smart Service (JSS) untuk memanfaatkan TIK.

Pemerintah kota Yogyakarta menerbitkan aplikasi Jogja Smart Service (JSS), tepat pada perayaan HUT Pemerintah kota Yogyakarta. Aplikasi ini merupakan bentuk perwujudan dari pengimplementasian e-government Pemerintah kota Yogyakarta. Ada beberapa peraturan e-government dalam tatanan kota Yogyakarta, yaitu penguatan infrastruktur jaringan, penguatan infrastruktur sistem informasi, integrasi data dan pengembangan aplikasi, data warehouse, dan kebijakan menuju smart city. Aplikasi JSS dapat digunakan sebagai sistem informasi pelayanan publik, misalnya seperti acara kreatif, informasi tingkat pendapatan daerah, informasi pengaduan, sistem pendidikan siswa online, kondisi infrastruktur, pariwisata, informasi tepat waktu, berita, dokumen perencanaan daerah, dan sebagainya. Aplikasi JSS memiliki layanan chatting bernama Jessica yang dapat digunakan untuk mengadukan berbagai kendala yang dihadapi masyarakat terhadap layanan yang tersedia di dalam aplikasi. Selanjutnya, penelitian ini akan memvisualisasikan data chat tersebut untuk mengetahui topik apa saja yang paling sering ditanyakan atau dikeluhkan masyarakat di kota Yogyakarta tentang aplikasi dan layanan JSS. Penelitian ini menggunakan Software Google Data Studio untuk visualisasi data agar hasilnya lebih mudah di pahami.

B. Identifikasi Masalah

Berdasarkan uraian latar belakang masalah tersebut, maka dapat diidentifikasi masalah sebagai berikut:

Praktik Kerja yang akan dilakukan di Dinas Komunikasi Informatika dan Persandian meliputi tahap analisa data dengan program komputer untuk mengetahui topik permasalahan yang sering diadukan pada layanan chat Jessica

sebagai bahan evaluasi sistem pada aplikasi JSS.

C. Batasan Masalah

Untuk menghindari perluasan masalah, maka batasan masalah pada kerja praktik ini adalah sebagai berikut:

1. Kerja Praktik pada Dinas Komunikasi Informatika dan Persandian Kota Yogyakarta dilaksanakan selama 3 bulan. Mulai dari bulan Oktober hingga Desember 2022.
2. Laporan KP diambil dari hasil analisis dataset layanan chat Jessica pada aplikasi Jogja Smart Service.
3. Visualisasi data pada layanan chat Jessica dilakukan dengan menggunakan Software Google Data Studio.

D. Rumusan Masalah

Berdasarkan pemaparan identifikasi masalah diatas, maka dapat dirumuskan sebagai berikut :

1. Bagaimana menggunakan Google Data Studio untuk visualisasi data topik permasalahan yang sering diadukan?
2. Bagaimana cara mengetahui topik permasalahan yang sering diadukan pada aplikasi Jogja Smart Service?

E. Tujuan Praktik Magang

Beberapa tujuan yang ingin dicapai dalam pelaksanaan program magang ini meliputi :

1. Mahasiswa mampu menganalisis dan menerapkan secara langsung teori/praktik yang didapatkan dari perkuliahan pada Kerja Praktik dengan mengolah dataset yang diberikan oleh Dinas Komunikasi Informatika dan Persandian Kota Yogyakarta.
2. Mendapatkan hasil visualisasi yang dapat digunakan sebagai bahan evaluasi sistem pada aplikasi JSS.

F. Manfaat Praktik Magang

Adapun manfaat yang dapat diperoleh dari kegiatan ini yaitu :

1. Bagi Mahasiswa
 - a. Mahasiswa dapat mengetahui mekanisme kerja pada instansi yang bersangkutan.
 - b. Mahasiswa terlatih untuk menganalisa permasalahan yang terjadi secara nyata.
 - c. Mahasiswa memperoleh ilmu dan pengalaman baru yang mungkin belum didapatkan sebelumnya.
2. Bagi Instansi
 - a. Memperoleh sumbangan pemikiran dan tenaga dalam rangka meningkatkan kinerja.
 - b. Memperoleh analisa terkait layanan chat Jessica pada aplikasi JSS yang dapat digunakan untuk evaluasi sistem aplikasi Jogja Smart Service.

BAB II

GAMBARAN INSTANSI

A. Umum (sejarah, visi misi, alamat, dan kontak)

1. Sejarah

Dinas Komunikasi, Informatika dan Persandian dibuat untuk membantu pengelolaan komunikasi dan enkripsi di dalam kota Yogyakarta. Dinas ini dibuat berdasarkan Peraturan Daerah Kota Yogyakarta Nomor 5 Tahun 2016 tentang Pembentukan dan Susunan Perangkat Daerah Kota Yogyakarta dan Peraturan Walikota Nomor 79 Tahun 2016 tentang Susunan organisasi, kedudukan, tugas, fungsi dan tata kerja Dinas Komunikasi, Informatika dan Persandian Kota Yogyakarta.

Pendirian Dinas Komunikasi, Informatika dan Persandian Kota Yogyakarta merupakan penggabungan dari Bagian Humas dan Informasi Sekretariat Daerah Kota Yogyakarta, Bagian Teknologi Informasi dan Telematika, dan pengelola sandi di Bagian Umum Sekretariat Daerah Kota Yogyakarta.

Menurut Perwal Nomor 79 Tahun 2016, Dinas Komunikasi, Informatika, dan Persandian wajib menyelenggarakan urusan pemerintahan daerah berdasarkan asas otonomi dan pengelolaan bersama di bidang komunikasi, teknologi informasi, persandian, dan statistik. Organisasi Diskominfo terdiri dari Kepala Dinas, Sub Bagian Tata Usaha, Bidang Informasi dan Statistik, Bidang Komunikasi Publik, Bidang Teknologi dan Informatika, dan Bidang Persandian dan Telekomunikasi.

Pada tahun 2017, Peraturan Walikota Yogyakarta nomor 81 menetapkan susunan dan organisasi Dinas Komunikasi Informatika dan Sandi Kota Yogyakarta. Perubahan ini berlaku untuk semua penyedia layanan di dalam kota, termasuk Layanan Kominfosandi.

Selanjutnya, pada Oktober 2021, berdasarkan Perda Kota Yogyakarta Nomor 4 Tahun 2020 akan dilakukan perubahan Perda Nomor 5 Tahun 2016 tentang Pembentukan dan Susunan Perangkat Daerah Kota Yogyakarta dan Peraturan Walikota Yogyakarta Nomor 105 Tahun 2021 tentang susunan, organisasi, kedudukan, tugas, fungsi dan tata kerja dinas komunikasi, Informatika dan Persandian Kota Yogyakarta.

Struktur organisasi Dinas Komunikasi Informatika dan Sandi didasarkan pada Perwal Nomor 105 Tahun 2021 yaitu Kepala Dinas, Sekretariat, Bidang Informasi dan Komunikasi Publik Bidang Sistem Informasi dan Statistik, Bidang Infrastruktur Telematika, dan Bidang Persandian dan Telekomunikasi.

2. Visi dan Misi

Visi : Terciptanya pelayanan publik yang terbaik melalui penerapan teknologi informatika dan komunikasi publik yang handal menuju Yogyakarta smart city

Misi :

1. Meningkatkan kualitas dan sebaran layanan diseminasi informasi melalui kemitraan dengan penyelenggara media dan mengikuti perkembangan teknologi informasi

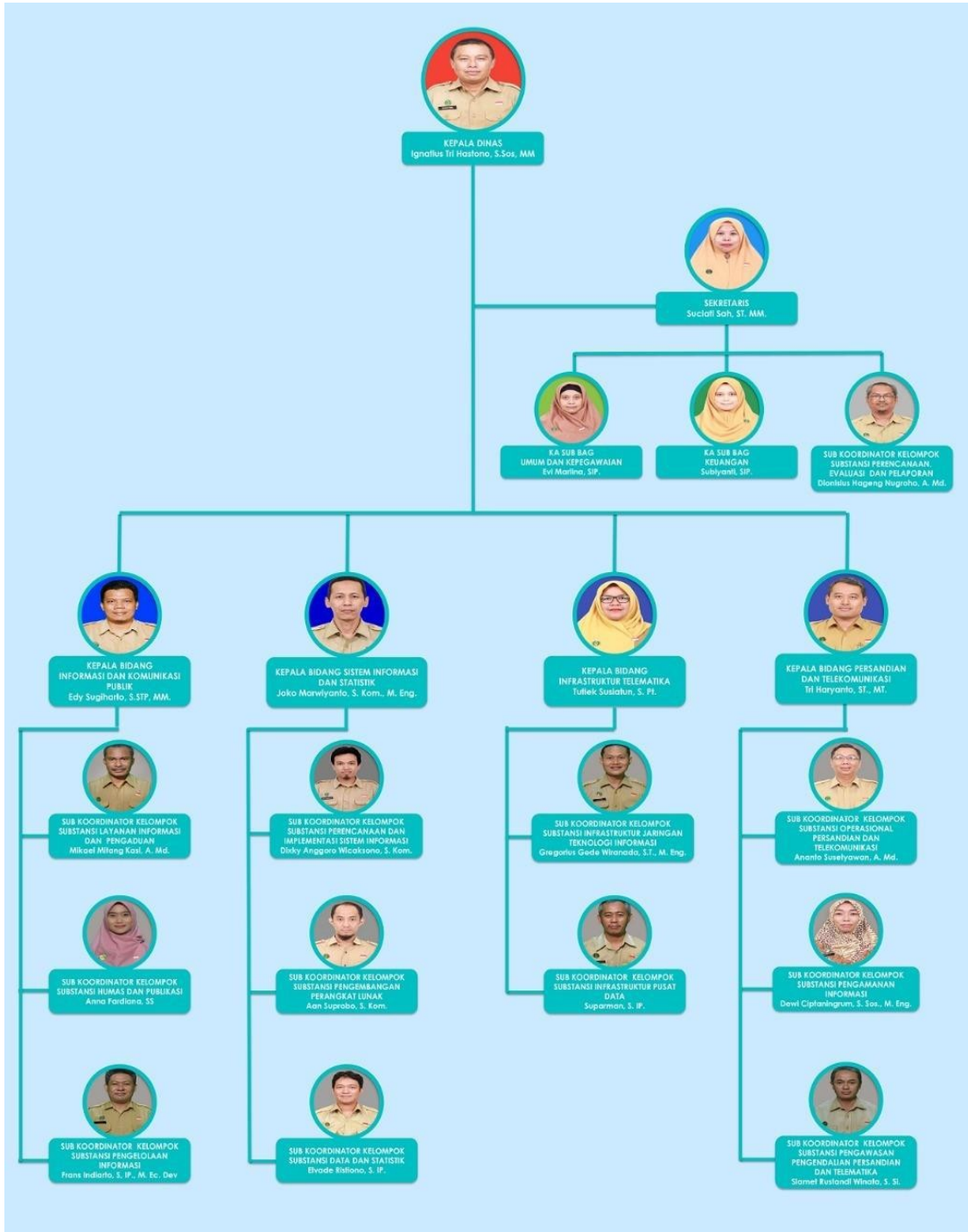
2. Memberdayakan potensi komunikasi di masyarakat dan peningkatan kerjasama dengan lembaga komunikasi dan informatika
3. Mendukung proses kerja internal pemerintah Kota Yogyakarta dengan pengembangan aplikasi dan sistem informasi manajemen yang handal, update, dan bisa menyesuaikan dengan kebutuhan masyarakat
5. Membangun kualitas jaringan teknologi informasi yang mandiri, terintegrasi, dan mampu bersinergi dengan pihak-pihak yang memiliki komitmen terhadap pengembangan smart city di Kota Yogyakarta
6. Menciptakan sistem kerja berbasis teknologi informasi yang handal dan terjamin keamanannya
7. Meningkatkan kualitas sumber daya manusia di bidang teknologi informasi dan komunikasi

3. Alamat dan kontak tempat magang

Dinas Komunikasi, Informatika dan Persandian Kota Yogyakarta terletak di Jl. Kenari No. 56 Yogyakarta. Dan Telp. (0274) 515865, 561270

B. Struktur Organisasi

Untuk mengetahui posisi struktur secara keseluruhan dari sudut pandang satu kesatuan divisi, berikut ini adalah struktur instansi Dinas Komunikasi, Informatika dan Persandian Kota Yogyakarta dapat dilihat pada Gambar 2.1.



Gambar 2. 1 Struktur Organisasi Instansi

C. Sumber Daya Manusia dan Sumber Daya Fisik

Sumber daya manusia yang ada pada Dinas Komunikasi Informatika dan Persandian beserta jabatannya dapat dilihat pada Tabel 2.1.

Tabel 2. 1 Struktur Organisasi Instansi

Nama	L/P	Jabatan
Ignatius Tri Hastoro, S.Sos., Mm.	L	Kepala Dinas
Suciati Sah, S.Sos., Mm.	P	Sekretaris
Evi Marlina, SIP.	P	KA SUB BAG Umum dan Kepegawaian
Subiyanti, SIP.	P	KA SUB BAG Keuangan
Dianisius Hageng Nugroho, A.Md.	L	SUB Koordinator Kelompok Substansi dan Perencanaan, Evaluasi, dan Pelaporan.
Egi Sugiharto, S.STP., MM.	L	Kepala Bidang Informasi dan Komunikasi Publik
Joko Marwoyanto, S.Kom., M.Eng.	L	Kepala Bidang Sistem Informasi dan Statistik
Tutik Susiatun, S.Pt.	P	Kepala Bidang Infrastruktur Telematika
Tri Haryanto, S.T., M.T.	L	Kepala Bidang Persandian dan Telekomunikasi
Mikael Milang Kasi, A.Md.	L	SUB Koordinator Kelompok Substansi Layanan Informasi dan Pengaduan
Anna Fardiana, SS.	P	SUB Koordinator Kelompok Substansi Humas dan Publikasi
Frans Indiarto, S.IP., M.Ec.Dev.	L	SUB Koordinator Kelompok Substansi Pengelolaan Informasi
Dixky Anggoro Wicaksono, S.Kom.	L	SUB Koordinator Kelompok Substansi Perencanaan dan Implementasi Sistem Informasi
Aan Suprobo, S.Kom.	L	SUB Koordinator Kelompok Substansi Pengembangan Perangkat Lunak
Elvade Ristiono, S.IP.	L	SUB Koordinator Kelompok Substansi Data dan Statistik
Gregorius Gede Wlranada, S.T., M.Eng.	L	SUB Koordinator Kelompok Substansi Infrastruktur Jaringan Teknologi Informasi
Suparman, S.IP.	L	SUB Koordinator Kelompok Substansi Infrastruktur Pusat Data
Ananto Susetyawan, A.Md.	P	SUB Koordinator Kelompok Substansi Operasional Persandian dan Telekomunikasi

Dewi Ciptaningrum, S.Sos., M.Eng.	P	SUB Koordinator Kelompok Substansi Pengamanan Informasi
Slamet Rustandi Winata, S.Si.	L	SUB Koordinator Kelompok Substansi Pengawasan Pengendalian Persandian dan Telematika

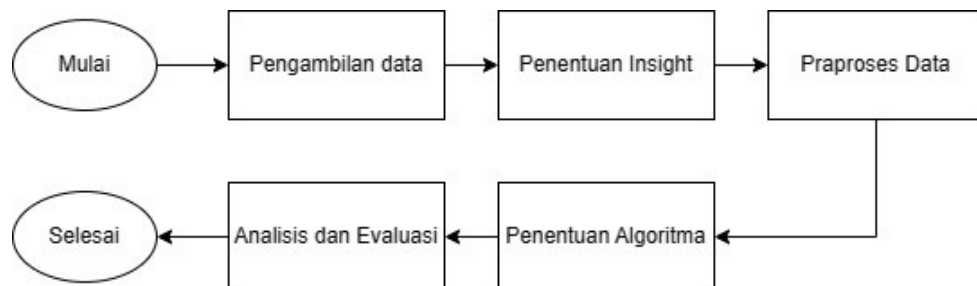
Sumber daya fisik yang ada pada Dinas Komunikasi dan Informatika dapat dilihat pada Tabel 2.2.

Tabel 2. 2 Sumber Daya Fisik DISKOMINFOSAN

Nama Barang	Jumlah	Satuan
Komputer	50	Unit
Ac	20	Unit
Meja	30	Unit
Kursi	30	Unit
Printer	10	Unit
CPU	50	Unit
Mouse	50	Unit
Keyboard	50	Unit
Lemari	20	Unit
Ruangan	15	Unit
Proyektor	5	Unit

D. Proses Bisnis

Proses Bisnis Kerja Praktik diawali dengan penentuan project didiskusikan bersama pembimbing KP. Project yang disepakati yaitu mengolah data chat Jessica pada aplikasi JSS pemerintah Kota Yogyakarta. Setelah memperoleh data Jessica, data tersebut diolah dan dianalisis sehingga mendapatkan kesimpulan berdasarkan insight yang diinginkan. Berikut rancangan dari proses bisnis yang akan dijalankan dapat dilihat pada Gambar 2.2.



Gambar 2. 2 Proses Bisnis

BAB III

TAHAPAN KEGIATAN PRAKTIK MAGANG

A. Lokasi Praktik Magang, Alamat, Kontak

1. Lokasi Kerja Praktik : Dinas Komunikasi Informatika dan Persandian
Kota Yogyakarta
2. Alamat : Jl. Kenari, Muja Muju, Kec. Umbulharjo, Kota
Yogyakarta, Daerah Istimewa Yogyakarta 55165
3. Kontak Pembimbing KP : 082137588899 (Elvade Ristiono, S.IP.)
4. Maps :

Berikut denah lokasi magang yaitu Dinas Komunikasi Informatika dan Persandian Kota Yogyakarta yang beralamat di Jl. Kenari, Muja Muju, Kec. Umbulharjo, Kota Yogyakarta, Daerah Istimewa Yogyakarta 55165 dapat dilihat pada Gambar 3.1.



Gambar 3.1 Maps Lokas Kerja Praktik

B. Rencana Observasi

Rencana observasi pada Kerja Praktek di Dinas Komunikasi Informatika dan Persandian Kota Yogyakarta adalah menganalisis dan visualisasi data untuk layanan chat Jessica pada aplikasi JSS (Jogja Smart Service) guna mengetahui topik yang paling banyak diadukan pada layanan chat tersebut.

C. Rancangan Jadwal Kegiatan Magang

Jadwal rencana kegiatan yang disusun setelah berdiskusi dengan tim kerja Praktik guna kelancaran kegiatan magang mahasiswa Universitas Ahmad Dahlan. Dalam Kerja Praktik di Dinas Komunikasi Informatika dan Persandian Kota Yogyakarta kegiatannya dilakukan selama 3 bulan yaitu dari bulan Oktober sampai Desember 2022. Pada minggu pelaksanaan ke-3 sampai dengan minggu ke-9 kosong karena menunggu hasil akhir dari pengelompokan topik menggunakan metode LDA. Rancangan jadwal kegiatan selama magang dapat dilihat pada Tabel 3.1 berikut:

Tabel 3. 1 Rencana dan Realisasi Kegiatan

No.	Nama Kegiatan	Minggu Pelaksanaan												Realisasi		
		1	2	3	4	5	6	7	8	9	10	11	12	Ya/Tidak	%	
1.	Penentuan Project	■													Ya	100
2.	Perolehan Data	■													Ya	100
3.	Studi Literatur	■	■												Ya	100
4.	Mendapatkan data hasil LDA										■				Ya	100
5.	Visualisasi Data											■			Ya	100
6.	Hasil												■		Ya	100
7.	Kesimpulan												■		Ya	100

BAB IV

HASIL PELAKSANAAN PRAKTIK MAGANG

A. Hasil Observasi Magang

1. Proses Bisnis

Proses awal dalam proses bisnis kerja praktik ini yaitu dilakukannya tahap *problem definition* yaitu suatu proses identifikasi yang harus dilakukan untuk menentukan masalah data layanan chat Jessica dari aplikasi Jogja Smart Service Center (JSS) pemerintah Kota Yogyakarta yang menjadi dasar perumusan pada penelitian ini. Tahap ini dilakukan dengan mendefinisikan masalah secara rinci untuk menentukan solusi pendekatan yang tepat. Selanjutnya, metode yang digunakan yaitu *Preprocessing* dan *Latent Dirichlet Allocation (LDA)* metode ini dipilih untuk menganalisis dokumen yang memiliki ukuran sangat besar. LDA bisa digunakan untuk meringkas, melakukan klusterisasi, menghubungkan atau memproses data yang sangat besar karena LDA menghasilkan daftar topik yang diberi bobot untuk masing-masing dokumen.

Tahap selanjutnya dilakukan visualisasi data dari topik LDA agar lebih mudah dibaca dan mengerti. Pada penyajian data menggunakan Google Data Studio diantaranya mempersiapkan sumber data, mempersiapkan koneksi antara sumber data dan Google Data Studio, dan menyajikan grafik yang dibutuhkan. Data yang digunakan yaitu data `log_messages` dimana data tersebut merupakan data chat yang keluar/balasan dari layanan chat Jessica aplikasi Jogja Service Center (JSS) pemerintah Kota Yogyakarta. Setelah menyelesaikan sumber data, selanjutnya mempersiapkan connector antara sumber data dengan Google Data Studio. Setelah tahapan pembuatan connector diselesaikan, maka masuk pada tahapan akhir yaitu membuat visualisasi. Dalam menyajikan visualisasi data ada berbagai jenis grafik dan tampilan yang tersedia, sehingga hasilnya bisa sesuai dengan bentuk yang diinginkan. Setelah visualisasi selesai maka akan mendapatkan hasil interpretasi yang akan disajikan ke pengguna. Berikut gambaran flowchart

proses bisnis dari kerja praktik untuk pengolahan data layanan chat Jessica.

2. Rincian Pekerjaan

Kerja Praktik di Dinas Komunikasi Informatika dan Persandian untuk mengolah data POSBINDU, dilaksanakan pada tanggal 3 Oktober 2022 sampai 30 Desember 2022. Namun untuk pemaparan akhir, dilakukan pada 16 Januari 2023. Tabel 4.1 merupakan rincian pekerjaan yang dilakukan saat kerja praktik di Dinas Komunikasi Informatika dan Persandian.

Tabel 4. 1 Rincian Pekerjaan

Hari/Tanggal	Jam Durasi	Kegiatan
Senin, 19 Sep 2022	1 Jam	Menyerahkan surat izin praktik magang di Dinas Kominfo Persandian Kota Yogyakarta.
Selasa, 20 Sep 2022	1 Jam	Penerimaan peserta magang oleh Dinas Kominfo Persandian Kota Yogyakarta.
Rabu, 21 Sep 2022	2 Jam	Pencarian topik project bersama tim.
Jumat, 23 Sep 2022	2 Jam	Rapat koordinasi internal kelompok substansi data dan statistik (Dinas Kominfosan Yogyakarta). Menentukan topik yang dipilih yaitu analisis dataset layanan chat Jessica pada aplikasi JSS.
Kamis, 06 Okt 2022	3 Jam	Mendapatkan dataset layanan chat Jessica pada aplikasi JSS.
Sabtu, 08 Okt 2022	4 Jam	Melakukan observasi pada data yang diterima dan melakukan studi literatur untuk menentukan metode yang akan diterapkan.
Rabu, 12 Okt 2022	4 Jam	Melaksanakan kegiatan sebagai panitia di BPS Kota Yogyakarta.
Sabtu, 15 Okt 2022	3 Jam	Menentukan metode dan menyiapkan ppt untuk paparan.

Kamis, 20 Okt 2022	1 Jam	Memaparkan progress berupa rencana tahapan pengolahan dataset yang diperoleh.
Rabu, 16 Des 2022	4 Jam	Melakukan proses visualisasi data.
Kamis, 17 Des 2022	1 Jam	Pengecekan kembali visualisasi data dan membuat ppt paparan akhir.
Senin, 19 Des 2022	3 Jam	Presentasi akhir bersama anggota dinas terkait program yang dilakukan

B. Pembahasan Magang

1. Problem yang Ditemukan di Tempat Magang

Permasalahan yang ditemukan di tempat magang sehingga diambil sebagai topik magang yaitu belum adanya analisis visualisasi data mengenai layanan chat Jessica pada aplikasi Jogja Smart Service. Maka dari itu dilakukan visualisasi data dengan menggunakan Google Data Studio untuk memudahkan dan mengetahui topik apa saja yang sering ditanyakan atau diadakan untuk dijadikan bahan evaluasi di tempat magang.

2. Analisis Terhadap Hasil Observasi

Berdasarkan hasil observasi serta diskusi dengan divisi bidang data dan statistik pada Diskominfo kota Yogyakarta, diketahui bahwa perlu adanya analisis visualisasi data pada layanan chat Jessica sehingga didapatkan topik atau aduan kendala yang sering dibahas pada aplikasi Jogja Smart Service (JSS) untuk dijadikan bahan evaluasi pada sistem aplikasi dan pemerintahan Kota Yogyakarta. Dengan begitu, penulis dapat menyusun analisa observasi menjadi laporan.

3. Capaian Magang

Berikut capaian dan hasil dari kegiatan magang yang sudah terlaksana dan ditampilkan dalam penjelasan dari proses bisnis dan analisa yang sudah didapatkan dari pengolahan data berdasarkan analisa pada hasil observasi.

a. Problem Definition

Pada tahap awal metode lifecycle of text mining ini yaitu mendefinisikan masalah (problem definition), penulis melakukan identifikasi masalah berdasarkan latar belakang penelitian. Studi kasus penelitian mengenai layanan chat jessica yang merupakan fasilitas yang terdapat dalam aplikasi Jogja Smart Service. Penulis mendefinisikan masalah pada text mining dengan topic modelling untuk mengetahui topik pembicaraan masyarakat Jogja tentang keluhan atau aduan menggunakan metode Preprocessing dan Latent Dirichlet Allocation (LDA). Data yang digunakan pada penelitian ini adalah dataset log_messages yang merupakan pesan yang masuk pada layanan chat Jessica pada aplikasi Jogja Smart Service.

b. Selecting Text Data Mining Approach

Pada tahap selecting text data mining approach, penulis menentukan pendekatan data mining yang digunakan pada penelitian ini mendapat masukan dari mentor kegiatan Kerja Praktik. Adapun masalah pada penelitian ini adalah menentukan Topic Modelling dengan Latent Dirichlet Allocation (LDA) dalam satu studi kasus chat tentang aduan masyarakat.

c. Data Collection

Data yang digunakan adalah dataset chat balasan dari layanan chat Jessica aplikasi JSS yang dioperasikan oleh Dinas Komunikasi Informatika dan Persandian Kota Yogyakarta. Data yang diambil adalah data chat yang masuk periode 26 April 2021 sampai dengan 03 Oktober 2022. Adapun tahap pengambilan data diantaranya data selection dan cleaning data. Untuk keperluan analisis, atribut yang dibutuhkan hanya

isi pesan dari chat balasan sehingga hanya atribut body yang akan digunakan. Setelah itu dilakukan cleaning data untuk menghilangkan missing value. Selain dataset chat, yang diperlukan pada penelitian ini yaitu kamus kata dasar KBBI, stopwords, dan kamus stopwords sastrawi.

d. Text Preprocessing

Tahap preprocessing dilakukan untuk menyeleksi data text agar menjadi lebih terstruktur dengan melalui serangkaian tahapan diantaranya:

1. Case Folding : Setelah melalui tahap data collection, selanjutnya tahap ini dilakukan untuk menyamaratakan semua karakter yang tersimpan yaitu dirubah menjadi lowercase atau huruf kecil.
2. Remove Punctuation : Setelah semua karakter huruf berubah menjadi kecil, maka pada tahap ini dilakukan penghapusan URL, @mention, #hashtag, emoticon, serta angka.
3. Tokenizing: Dalam tahap ini dilakukan pemisahan kata pada setiap kalimat dengan tujuan untuk proses analisis teks lebih lanjut.
4. Stopword: Proses stopwords dilakukan untuk mengurangi jumlah kata yang tidak penting. Daftar kata tidak penting (stoplist) yang digunakan adalah database kamus stopword sastrawi dan stopword manual.

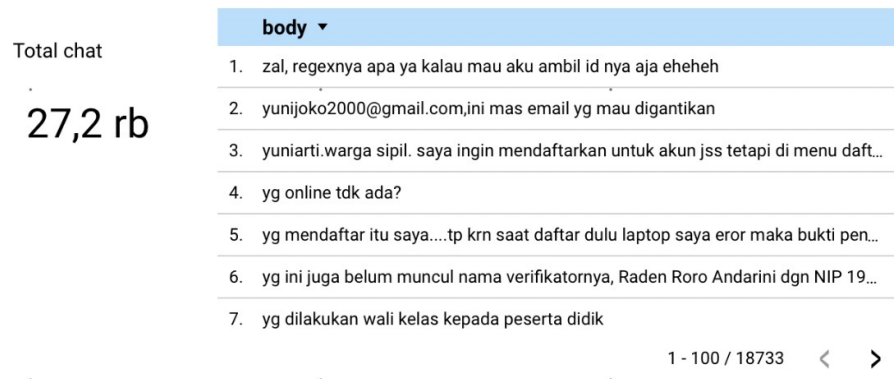
e. Analysis

1. Lantent Dirichlet Allocation (LDA)

Hasil dari analisis pemodelan topik dengan metode Latent Dirichlet Allocation (LDA) pada data pendapat masyarakat untuk pengelompokan topik chat bot jessica pada aplikasi jss diperoleh jumlah topik sebanyak 6 dengan coherence score sebesar 0.5644.

2. Visualisasi Data

Analisis tampilan visualisasi data dari topik permasalahan yang serin diadukan pada chat Jessica, dengan total sebanyak 6 topik yang divisualisasi agar lebih mudah di mengerti.



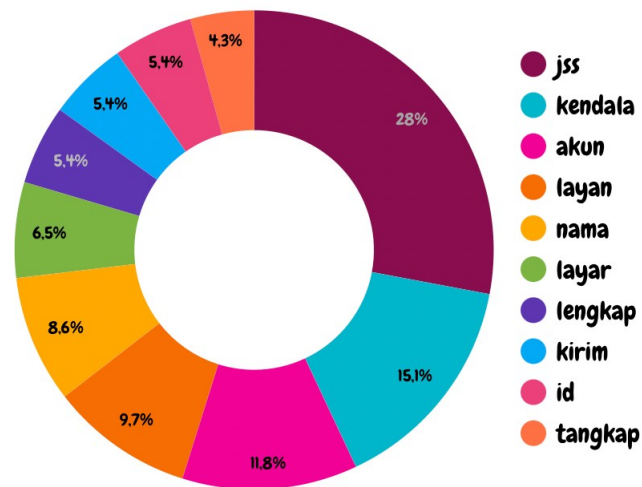
Gambar 4.1 Jumlah Chat Atribut body

Pada gambar 4.1 terdapat Jumlah chat atribut *body* pada layanan chat Jessica dari tanggal 26 April 2021 sampai 03 Oktober 2022 sebanyak 27.200 chat. Dilakukan analisis visualisasi data dari hasil pemodelan topik chatbot Jessica dengan Google Data Studio. Pada visualisasi data ini menggunakan grafik pie atau grafik lingkaran karena akan menampilkan perbandingan antar topik dari hasil LDA yang didapat agar lebih mudah dibaca. Adapun keunggulan menggunakan grafik pie yaitu menghemat space dan secara langsung dapat memperlihatkan perbandingan atau frekuensi masing-masing ukuran data melalui besar sektornya. Berikut penjelasan visualisasi output dari penelitian ini.

1) Visualisasi LDA Topik 1

Berdasarkan hasil perolehan data pada pemodelan topik LDA maka hasil tersebut divisualisasikan dan ditampilkan dalam bentuk grafik pie guna untuk memperlihatkan perbandingan masing- masing ukuran data dengan menggunakan Google Data Studio, visualisasi topik 1 dapat dilihat pada Gambar 4.2

VISUALISASI DATA HASIL LDA TOPIK 1



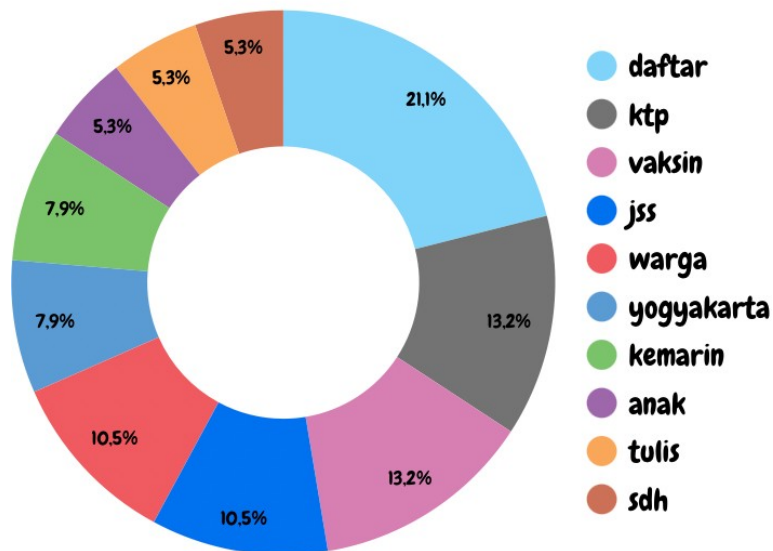
Gambar 4.2 Visualisasi Topik 1

Berdasarkan tampilan hasil visualisasi Topik 1 didapatkan hasil kata-kata yang sering muncul dan berkaitan satu sama lain adalah jss, kendala, akun, layer, lengkap, jessica, error, kirim, id, dan password. Kesimpulan dari visualisasi topik 1 diatas, yaitu adanya error pada sistem login akun JSS menggunakan id dan password yang sering dialami oleh pengguna. Sehingga berdasarkan beberapa data yang diperoleh dapat disimpulkan bahwa pada visualisasi topik 1 membahas tentang kendala penggunaan atau panduan pada jessica, serta membicarakan tentang cara masuk kedalam layanan JSS karena mengalami error.

2) Visualisasi LDA Topik 2

Berdasarkan hasil perolehan data pada pemodelan topik LDA maka hasil tersebut divisualisasikan dan ditampilkan dalam bentuk grafik pie guna untuk memperlihatkan perbandingan masing-masing ukuran data dengan menggunakan Google Data Studio, visualisasi topik 2 dapat dilihat pada Gambar 4.3.

VISUALISASI DATA HASIL LDA TOPIK 2



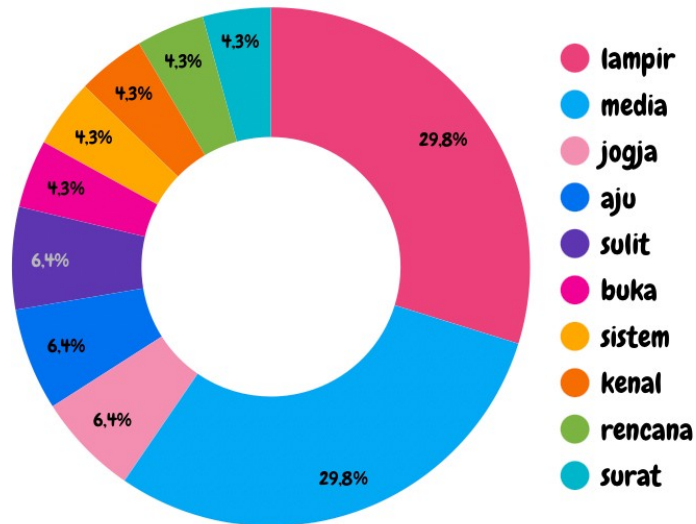
Gambar 4.3 Visualisasi Topik 2

Berdasarkan tampilan hasil visualisasi pada Topik 2 didapatkan hasil kata-kata yang sering muncul dan berkaitan satu sama lain adalah daftar, vaksin, ktp, jss, nik, dan lainnya. Kesimpulan dari visualisasi topik 2 diatas, yaitu pada pendaftaran vaksin masyarakat akan menggunakan akun JSS dimana diwajibkan untuk melakukan pendaftaran menggunakan KTP dan NIK yang digunakan untuk mendaftar vaksin.

3) Visualisasi LDA Topik ke 3

Berdasarkan hasil perolehan data pada pemodelan topik LDA maka hasil tersebut divisualisasikan dan ditampilkan dalam bentuk grafik pie guna untuk memperlihatkan perbandingan masing- masing ukuran data dengan menggunakan Google Data Studio, visualisasi topik 3 dapat dilihat pada Gambar 4.4.

VISUALISASI DATA HASIL LDA TOPIK 3



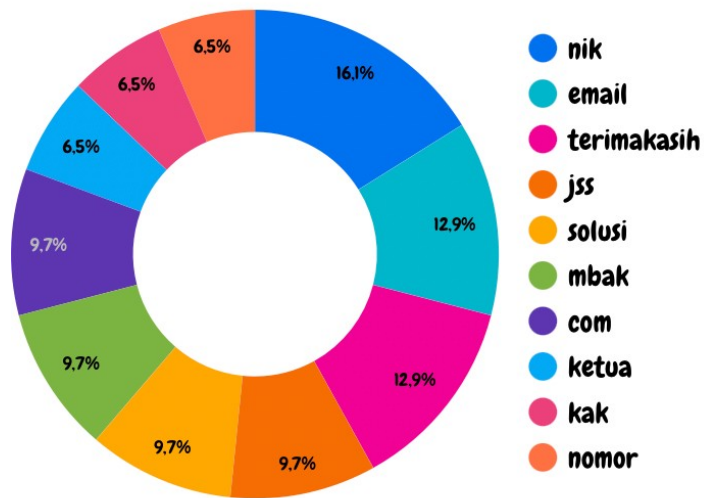
Gambar 4.4 Visualisasi Topik 3

Berdasarkan tampilan hasil visualisasi pada Topik 3 didapatkan hasil kata-kata yang sering muncul dan berkaitan satu sama lain adalah melampirkan media. Pada topik ini pengirim pesan melakukan aksi melampirkan media pada chat yang terdiri gambar, chipertext, dokumen, dan video.

4) Visualisasi LDA Topik 4

Berdasarkan hasil perolehan data pada pemodelan topik LDA maka hasil tersebut divisualisasikan dan ditampilkan dalam bentuk grafik pie guna untuk memperlihatkan perbandingan masing- masing ukuran data dengan menggunakan Google Data Studio, visualisasi topik 4 dapat dilihat pada Gambar 4.5.

VISUALISASI DATA HASIL LDA TOPIK 4



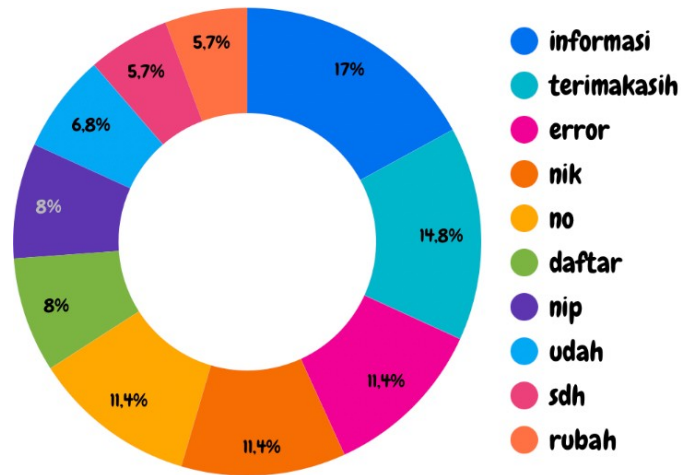
Gambar 4.5 Visualisasi Topik 4

Berdasarkan tampilan hasil visualisasi pada Topik 4 didapatkan hasil kata-kata yang sering muncul dan berkaitan satu sama lain adalah nik, email, JSS, ketua, nomor, dan lainnya. Kesimpulan dari visualisasi topik 4 diatas, yaitu dalam mengubah status user sebagai ketua RT terdapat berbagai kendala yang ditemui yaitu sulitnya mengubah status ketua RT sebagai user yang sebelumnya merupakan warga sehingga tidak dapat melakukan aksi akses JSS sebagai user. Selain itu, pada informasi melalui email terdapat kendala yang ditemui dalam aplikasi yaitu saat masyarakat mengajukan permohonan dalam aplikasi tersebut tidak ada pemberitahuan lebih lanjut.

5) Visualisasi LDA Topik 5

Berdasarkan hasil perolehan data pada pemodelan topik LDA maka hasil tersebut divisualisasikan dan ditampilkan dalam bentuk grafik pie guna untuk memperlihatkan perbandingan masing-masing ukuran data dengan menggunakan Google Data Studio, visualisasi topik 5 dapat dilihat pada Gambar 4.6.

VISUALISASI DATA HASIL LDA TOPIK 5



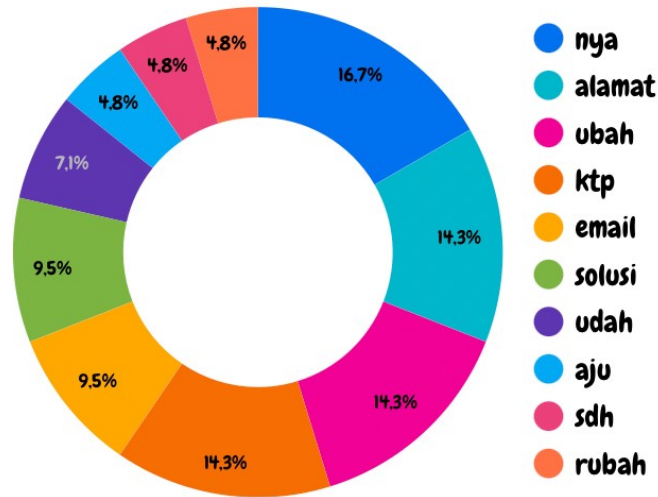
Gambar 4.6 Visualisasi Topik 5

Berdasarkan tampilan hasil visualisasi pada Topik 5 didapatkan hasil kata-kata yang sering muncul dengan bobot tertinggi adalah informasi. Pada aplikasi JSS merupakan penghubung masyarakat dengan berbagai informasi yang ada di Jogjakarta. Berbagai kendala yang ditemui oleh masyarakat dapat di pecahkan dengan aplikasi JSS, sehingga masyarakat dapat menanyakan terkait berbagai informasi yang dibutuhkan.

6) Model LDA Topik 6

Berdasarkan hasil perolehan data pada pemodelan topik LDA maka hasil tersebut divisualisasikan dan ditampilkan dalam bentuk grafik pie guna untuk memperlihatkan perbandingan masing- masing ukuran data dengan menggunakan Google Data Studio, visualisasi topik 6 dapat dilihat pada Gambar 4.7.

VISUALISASI DATA HASIL LDA TOPIK 6



Gambar 4.7 Visualisasi Topik 6

Berdasarkan tampilan hasil visualisasi pada Topik 6 didapatkan hasil kata-kata yang sering muncul dan berkaitan satu sama lain adalah ktp, alamat, ubah, aju, dan lainnya. Kesimpulan dari visualisasi topik 4 diatas, yaitu mengenai pengajuan permohonan di aplikasi JSS. Pada JSS masyarakat dapat melakukan pengajuan, beberapa pengajuan tersebut adalah permohonan KIA, rekam KTP EL, pindah domisili, dan lainnya.

4. Keberlanjutan Magang

Dalam kegiatan praktik magang ini penulis tidak melakukan keberlanjutan dengan Dinas Komunikasi Informatika dan Persandian Kota Yogyakarta, sehingga penulis hanya melaksanakan kegiatan praktik magang, memperoleh ilmu, dan relasi atau koneksi.

BAB V

PENUTUP

A. Kesimpulan

Setelah pelaksanaan Kerja Praktik (KP) yang dilaksanakan selama 3 (tiga) bulan di Dinas Komunikasi Informatika dan Persandian Kota Yogyakarta, analisis yang dilakukan yaitu hasil visualisasi data menggunakan Google Data Studio berdasarkan topik permasalahan yang sering diadukan untuk layanan chat Jessica pada aplikasi Jogja Smart Service. Tahap Penelitian sebelumnya adalah mengimplementasikan algoritma LDA dengan keluaran berupa topik yang paling banyak muncul dalam layanan chat Jessica. Ada 6 topik yang didapatkan dalam pengimplementasian LDA dan dilakukannya visualisasi Data menggunakan Google Data Studio. Hasil dari visualisasi data yaitu tampilan perbandingan dari topik-topik permasalahan yang sering diadukan pada chat Jessica, dengan total sebanyak 6 topik yang divisualisasi agar lebih mudah dimengerti.

B. Saran

Adapun saran yang dapat disimpulkan terkait pengerjaan proyek pada Kerja Praktik ini antara lain:

1. Melakukan perbaikan pada sistem aplikasi Jogja Smart Service (JSS) mengingat adanya kendala-kendala atau error pada saat login dan menggunakan aplikasi.
2. Menyediakan sistem chat untuk pertanyaan yang sering ditanyakan atau Frequently Asked Questions (FAQ) agar admin tidak membalas pertanyaan yang sama secara berulang.
3. Analisis ini diharapkan bisa memberikan informasi yang diinginkan serta dapat bermanfaat bagi pengguna dan pengelola aplikasi Jogja Smart Service sebagai bahan evaluasi sistem kedepannya agar lebih baik.

LAMPIRAN

A. Surat Izin Praktik Magang



UNIVERSITAS AHMAD DAHLAN FAKULTAS TEKNOLOGI INDUSTRI

KAMPUS 4 : Jl. Ahmad Yani, Tamanan, Banguntapan, Bantul, Yogyakarta. 55191
☎ 0274 563515, 511830 ext 41211 📠 0819 9079 0454 🌐 fti.uad.ac.id ✉ fti@uad.ac.id

Nomor : F2.1/133/D.33/IX/2022

19 September 2022

Hal : Permohonan Izin Kerja Praktek

Kepada

Yth. Kepala

Dinas Komunikasi Informatika dan Persandian Kota Yogyakarta

Jl. Kenari

Muja-Muja,, Umbulharjo, Yogyakarta, Daerah Istimewa Yogyakarta 55165

Dengan hormat,

Dalam rangka memberikan kesempatan kepada mahasiswa dalam mengimplementasikan disiplin ilmu yang diperoleh di bangku kuliah dengan keadaan sebenarnya, Pimpinan Fakultas Teknologi Industri mengajukan permohonan kepada Bapak/Ibu untuk berkenan memberikan ijin kepada mahasiswa kami tersebut di bawah ini untuk melaksanakan **Kerja Praktek** di instansi yang Bapak/Ibu pimpin.

Adapun data mahasiswa sebagai berikut:

No	Nama	NIM
1.	Indah Septiani	1900018291
2.	Nadya Cantika Putri Abdun	1900018288
3.	Maya Sofiya Intantri	1900018298
4.	-	-

Program Studi : **S1 Informatika**

Fakultas : Teknologi Industri

Sebagai bahan pertimbangan Bapak/Ibu kami sampaikan bahwa sebagai konsekuensinya, mahasiswa yang bersangkutan bersedia memenuhi persyaratan administratif yang diperlukan.

Waktu Pelaksanaan Kerja Praktek Mulai Tanggal 03 Oktober - 30 Desember 2022.

Atas perhatian dan kerjasama Bapak/Ibu, kami sampaikan terima kasih.



Hormat kami,
a.n Dekan
Wakil Dekan

Sri Winiarti, S.T., M.Cs.
NIY. 60020388

B. Surat Keterangan Telah Menyelesaikan Praktik Magang



Yogyakarta, 31 Januari 2023

Nomor : 423/0224
Sifat : Biasa
Lampiran :
Hal : Surat Keterangan Selesai Kerja Praktik a.n
Maya Sofiya Intantri, Nadya Cantika Putri
Abdun, dan Indah Septiani

Kepada
Yth. Dekan Fakultas Teknologi
Industri Universitas Ahmad
Dahlan Yogyakarta

di
Yogyakarta

Dengan hormat,

Memperhatikan surat dari Fakultas Teknologi Industri Universitas Ahmad Dahlan Yogyakarta Nomor : F2.1/142/D.33/X/2022 tanggal 3 Oktober 2022 Perihal : Permohonan Izin Kerja Praktek, bersama ini kami sampaikan bahwa mahasiswa sebagai berikut:

No	Nama	NIM	Program Studi
1	Maya Sofiya Intantri	1900018298	S1 Informatika
2	Nadya Cantika Putri Abdun	1900018288	S1 Informatika
3	Indah Septiani	1900018291	S1 Informatika

Bersama ini kami sampaikan bahwa mahasiswa tersebut telah melaksanakan Kerja Praktik di Dinas Komunikasi Informatika Dan Persandian Kota Yogyakarta pada tanggal 03 Oktober - 30 Desember 2022

Demikian surat keterangan ini dibuat, untuk dipergunakan sebagaimana mestinya

Kepala Dinas

IGNATIUS TRIHASTONO, S.Sos., M.M.
NIP. 196907231996031005



SEGORO AMARTO
SEMANGAT GOTONG ROYONG AGAWE MAJUNE NGAYOGYAKARTA
KEMANDIRIAN – KEDIPLINAN – KEPEDULIAN- KEBERSAMAAN

C. Logbook

**LOG BOOK KERJA PRAKTIK MAHASISWA PROGRAM
STUDI S1 INFORMATIKA T.A 2022 / 2023
(WAJIB DIISI DAN MASUK DALAM PENILAIAN)**

Nim : 1900018288
 Nama Mahasiswa : Nadya Cantika Putri Abdun
 Judul Praktik Magang : Analisis Chatbot Jessica pada Aplikasi JSS
 Dosen Pembimbing : Sri Winiarti, S.T., M.Cs
 Pembimbing Lapangan : Elvade Ristono, S.I.P.

Petunjuk Pengisian Log Book

1. Log book di isi per minggu
2. Log book ditulis tangan
3. Setiap kegiatan di paraf oleh pembimbing lapangan/ dosen pembimbing Praktik Magang
4. Log book per minggu di paraf oleh dosen pengampu kelas Praktik Magang
5. Jumlah bimbingan minimal 4x

Logbook Minggu 4 sd 7 (sebelum UTS)

No	Kegiatan dan Lokasi Praktik Magang	Waktu Pelaksanaan		Hasil	Kendala, Rencana Perubahan (Jika ada)	Paraf Pembimbing Lapangan	Paraf Dosen Pembimbing Praktik Magang
		Hari/TGL	Jam Durasi				
1.	Penentuan topik magang	Jumat, 23 September 2022	1 Jam	Berhasil menentukan topik yang dipilih			
2.	Mendapatkan dataset dari instansi	Kamis, 06 Oktober 2022	1 Jam	Berhasil mendapatkan dataset			
3.	Kegiatan FGD Metadata dengan BPS Kota Yogyakarta	Rabu, 12 Oktober 2022	4 Jam	Berhasil melaksanakan kegiatan			
4.	Paparan progress proyek magang	Kamis, 20 Oktober 2022	3 Jam	Berhasil memaparkan progress			
5.	Paparan progress proyek magang	Kamis, 17 November 2022	1 Jam	Berhasil memaparkan progress			

Catatan Pembimbing Lapangan/Dosen Pembimbing Praktik Magang / Dosen Pengampu Kelas Praktik Magang:

.....

Dosen Pengampu Kelas Praktik Magang

(Nuri Anwar, S.T., M.Kom)

Yogyakarta, 19 November 2022

Mahasiswa

(Nadya Cantika Putri Abdun)





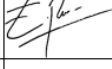

LOG BOOK KERJA PRAKTEK MAHASISWA
PROGRAM STUDI S1 INFORMATIKA, UAD T.A 2022 / 2023
(WAJIB DIISI DAN MASUK DALAM PENILAIAN)

Nim : 1900018288
Nama Mahasiswa : Nadya Cantika Putri Abdun
Judul Praktik Magang : Analisis Chatbot Jessica pada Aplikasi JSS
Dosen Pembimbing : Sri Winiarti, S.T., M.Cs
Pembimbing Lapangan : Elvade Ristiono, S.I.P.

Petunjuk Pengisian Log Book

1. Log book di isi per minggu
2. Log book ditulis tangan
3. Setiap kegiatan di paraf oleh pembimbing lapangan/ dosen pembimbing Praktik Magang
4. Log book per minggu di paraf oleh dosen pengampu kelas Praktik Magang
5. Jumlah bimbingan minimal 3x

Logbook Minggu 8 sd 10 (setelah UTS)

No	Kegiatan dan Lokasi Praktik Magang	Waktu Pelaksanaan		Hasil	Kendala, Rencana Perubahan (Jika ada)	Paraf Pembimbing Lapangan	Paraf Dosen Pembimbing Praktik Magang
		Hari/TGL	Jam Durasi				
1.	Paparan lanjutan progress proyek magang	Rabu, 23 November 2022	1 Jam	Berhasil memaparkan update progress magang			
2.	Paparan lanjutan progress proyek magang	Kamis, 15 Desember 2022	1 Jam	Berhasil memaparkan update progress magang			
3.	Paparan hasil akhir proyek magang	Senin, 16 Januari 2023	3 Jam	Berhasil memaparkan hasil akhir proyek magang			

Catatan Pembimbing Lapangan/Dosen Pembimbing Praktik Magang / Dosen Pengampu Kelas Praktik Magang:

.....

.....

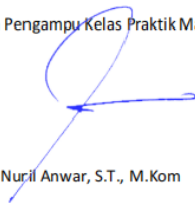
.....

.....

.....

Dosen Pengampu Kelas Praktik Magang

(Nuril Anwar, S.T., M.Kom)



Yogyakarta, 16 Januari 2023

Mahasiswa

(Nadya Cantika Putri Abdun)



D. Dokumentasi Kegiatan Magang



E. Source Code

ANALISIS CHATBOT JESSICA PADA APLIKASI JSS MENGGUNAKAN ALGORITMA LATENT DIRICHLET ALLOCATION

Instalasi beberapa library yang diperlukan

```
!pip install --upgrade gensim
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Requirement already satisfied: gensim in /usr/local/lib/python3.8/dist-packages (4.3.0)
Requirement already satisfied: numpy>=1.18.5 in /usr/local/lib/python3.8/dist-packages (from gensim) (1.22.4)
Requirement already satisfied: scipy>=1.7.0 in /usr/local/lib/python3.8/dist-packages (from gensim) (1.7.3)
Requirement already satisfied: FuzzyTM>=0.4.0 in /usr/local/lib/python3.8/dist-packages (from gensim) (2.0.5)
Requirement already satisfied: smart-open>=1.8.1 in /usr/local/lib/python3.8/dist-packages (from gensim) (6.3.0)
Requirement already satisfied: pyfume in /usr/local/lib/python3.8/dist-packages (from FuzzyTM>=0.4.0->gensim) (0.2.25)
Requirement already satisfied: pandas in /usr/local/lib/python3.8/dist-packages (from FuzzyTM>=0.4.0->gensim) (1.3.5)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.8/dist-packages (from pandas->FuzzyTM>=0.4.0)
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.8/dist-packages (from pandas->FuzzyTM>=0.4.0->ger)
Requirement already satisfied: simpful in /usr/local/lib/python3.8/dist-packages (from pyfume->FuzzyTM>=0.4.0->gensim)
Requirement already satisfied: fst-psy in /usr/local/lib/python3.8/dist-packages (from pyfume->FuzzyTM>=0.4.0->gensim)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.8/dist-packages (from python-dateutil>=2.7.3->pandas)
Requirement already satisfied: miniful in /usr/local/lib/python3.8/dist-packages (from fst-psy->pyfume->FuzzyTM>=0.4.0)
Requirement already satisfied: requests in /usr/local/lib/python3.8/dist-packages (from simpful->pyfume->FuzzyTM>=0.4.0)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.8/dist-packages (from requests->simpful->pyfume->FuzzyTM>=0.4.0)
Requirement already satisfied: urllib3<1.27,>=1.21.1 in /usr/local/lib/python3.8/dist-packages (from requests->simpful->pyfume->FuzzyTM>=0.4.0)
Requirement already satisfied: certifi>=2017.4.17 in /usr/local/lib/python3.8/dist-packages (from requests->simpful->pyfume->FuzzyTM>=0.4.0)
Requirement already satisfied: chardet<5,>=3.0.2 in /usr/local/lib/python3.8/dist-packages (from requests->simpful->pyfume->FuzzyTM>=0.4.0)
```

```
#pip install --upgrade numpy
```

```
import gensim
```

```
pip install pyldavis==3.2.1
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting pyldavis==3.2.1
  Downloading pyLDavis-3.2.1.tar.gz (1.7 MB)
    1.7/1.7 MB 23.4 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: wheel>=0.23.0 in /usr/local/lib/python3.8/dist-packages (from pyldavis==3.2.1) (0.38.4)
Requirement already satisfied: numpy>=1.9.2 in /usr/local/lib/python3.8/dist-packages (from pyldavis==3.2.1) (1.22.4)
Requirement already satisfied: scipy>=0.18.0 in /usr/local/lib/python3.8/dist-packages (from pyldavis==3.2.1) (1.7.3)
Requirement already satisfied: joblib>=0.8.4 in /usr/local/lib/python3.8/dist-packages (from pyldavis==3.2.1) (1.2.0)
Requirement already satisfied: Jinja2>=2.7.2 in /usr/local/lib/python3.8/dist-packages (from pyldavis==3.2.1) (2.11.3)
Requirement already satisfied: NumExpr in /usr/local/lib/python3.8/dist-packages (from pyldavis==3.2.1) (2.8.4)
Requirement already satisfied: future in /usr/local/lib/python3.8/dist-packages (from pyldavis==3.2.1) (0.16.0)
Collecting funcy
  Downloading funcy-1.18-py2.py3-none-any.whl (33 kB)
Requirement already satisfied: pandas>=0.17.0 in /usr/local/lib/python3.8/dist-packages (from pyldavis==3.2.1) (1.3.5)
Requirement already satisfied: MarkupSafe>=0.23 in /usr/local/lib/python3.8/dist-packages (from Jinja2>=2.7.2->pyldavis) (2.0.1)
Requirement already satisfied: python-dateutil>=2.7.3 in /usr/local/lib/python3.8/dist-packages (from pandas>=0.17.0->pyldavis) (2.8.2)
Requirement already satisfied: pytz>=2017.3 in /usr/local/lib/python3.8/dist-packages (from pandas>=0.17.0->pyldavis) (2017.3)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.8/dist-packages (from python-dateutil>=2.7.3->pandas) (1.16.0)
Building wheels for collected packages: pyldavis
  Building wheel for pyldavis (setup.py) ... done
  Created wheel for pyldavis: filename=pyLDavis-3.2.1-py2.py3-none-any.whl size=136188 sha256=a3a3a79a4af2056f8e550b00c1e0d3a914105d57dd
  Stored in directory: /root/.cache/pip/wheels/ae/5f/79/278ffb79ffde795f9686b558a8e313971ee0d3a914105d57dd
Successfully built pyldavis
Installing collected packages: funcy, pyldavis
Successfully installed funcy-1.18 pyldavis-3.2.1
```

```
#kosakata Indonesia
```

```
!pip install Sastrawi
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting Sastrawi
  Downloading Sastrawi-1.0.1-py2.py3-none-any.whl (209 kB)
    209.7/209.7 KB 6.3 MB/s eta 0:00:00
Installing collected packages: Sastrawi
Successfully installed Sastrawi-1.0.1
```

```
pip install swifter
```

```
Looking in indexes: https://pypi.org/simple, https://us-python.pkg.dev/colab-wheels/public/simple/
Collecting swifter
  Downloading swifter-1.3.4.tar.gz (830 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 830.9/830.9 KB 14.6 MB/s eta 0:00:00
  Preparing metadata (setup.py) ... done
Requirement already satisfied: pandas>=1.0.0 in /usr/local/lib/python3.8/dist-packages (from swifter) (1.3.5)
Collecting psutil>=5.6.6
  Downloading psutil-5.9.4-cp36-abi3-manylinux_2_12_x86_64.manylinux2010_x86_64.manylinux_2_17_x86_64.manylinux2014_x86_64.whl (280.2/280.2 KB 26.7 MB/s eta 0:00:00)
Requirement already satisfied: dask[dataframe]>=2.10.0 in /usr/local/lib/python3.8/dist-packages (from swifter) (2022.12.0)
Requirement already satisfied: tqdm>=4.33.0 in /usr/local/lib/python3.8/dist-packages (from swifter) (4.64.1)
Requirement already satisfied: ipywidgets>=7.0.0 in /usr/local/lib/python3.8/dist-packages (from swifter) (7.7.1)
Requirement already satisfied: cloudpickle>=0.2.2 in /usr/local/lib/python3.8/dist-packages (from swifter) (2.2.0)
Requirement already satisfied: parso>0.4.0 in /usr/local/lib/python3.8/dist-packages (from swifter) (0.8.3)
Requirement already satisfied: bleach>=3.1.1 in /usr/local/lib/python3.8/dist-packages (from swifter) (5.0.1)
Requirement already satisfied: six>=1.9.0 in /usr/local/lib/python3.8/dist-packages (from bleach>=3.1.1->swifter) (1.16.0)
Requirement already satisfied: webencodings in /usr/local/lib/python3.8/dist-packages (from bleach>=3.1.1->swifter) (0.11.0)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.8/dist-packages (from dask[dataframe]>=2.10.0->cloudpickle)
```

▼ Preparing Data dan Cleaning

```
Requirement already satisfied: ipynb2pdf in /usr/local/lib/python3.8/dist-packages (from notebook>=4.4.1->widgets) 0.10.0
import pandas as pd
import numpy as np
import nltk
```

```
Requirement already satisfied: Send2Trash in /usr/local/lib/python3.8/dist-packages (from notebook>=4.4.1->widgets) 1.8.0
#import data dari GDrive
```

```
! gdown --id 13EDy-koYKccrxU0ZhfzIQJ3RSSQ_64
```

```
/usr/local/lib/python3.8/dist-packages/gdown/cli.py:127: FutureWarning: Option `--id` was deprecated in version 4.3.1 and will be removed in a future version. Use `--url` instead.
  warnings.warn(
Downloading...
From: https://drive.google.com/uc?id=13EDy-koYKccrxU0ZhfzIQJ3RSSQ\_64
To: /content/dataBerita.xlsx
100% 4.65M/4.65M [00:00<00:00, 229MB/s]
```

```
#menggunakan dataframe dataSB
```

```
dataSB = pd.read_excel('/content/dataBerita.xlsx', sheet_name="log_messages") #lokasi file
```

```
dataSB.head()
```

	_id	ack	author	body	broadcast	created_at	description	deviceType	forwardingScore
0	6086e3941b141e02714a934a	-1.0	NaN	Keluar sprti ini,bgmn solusinya	False	2021-04-26 16:00:20	NaN	NaN	NaN 62
1	60875a061b141e02714a937b	-1.0	NaN	Selamat pagi pak/bu mohon maaf mau bertanya. S...	False	2021-04-27 00:25:42	NaN	NaN	NaN 6
2	60875f811b141e02714a9382	-1.0	NaN	Selamat pagi, terkait dengan fitur e kinerja s...	False	2021-04-27 00:49:05	NaN	NaN	NaN 6
				Salam Sehat, ...					

▼ Data selection

```
menghapus atribut yang tidak diperlukan
```

```
dataSB = dataSB.drop(['_id', 'ack', 'author', 'body', 'broadcast', 'created_at', 'description', 'deviceType', 'forwardingScore', 'from', 'fromMe', 'hasMedia'])
```

```
dataSB = dataSB.drop(['_id', 'ack', 'author', 'broadcast', 'created_at', 'description', 'deviceType', 'forwardingScore', 'from', 'fromMe', 'hasMedia'])
dataSB.head()
```



```

body
22856  solusi nya bagaimana supaya bisa pakai email ...
20405  Siapp terimakasih banyak admin 🙏
23076  baik, terima kasih atas informasinya
6747  Salam Sehat, JSS Support di sini. 🙏 😊 \nBila ad...

# dataSB.info()

# dataSB.isnull()

# dataSB.isna().sum()

#menampilkan missing value pada atribut

#dataSB.isnull().sum()

#mengisi missing value pada atribut body

dataSB['body'].fillna('melampirkan media', inplace=True)

dataSB.isnull().sum()

body          0
body_tokens   0
body_tokens_fdist  0
body_tokens_WSW  0
body_normalized  0
body_tokens_stemmed  0
body_tokens_stemmed2  0
dtype: int64

```

▼ Case folding

Mengubah menjadi huruf kecil

```

# ----- Case Folding -----
# menggunakan fungsi Series.str.lower() pada Pandas

dataSB['body'] = dataSB['body'].str.lower()

print('Case Folding Result : \n')
print(dataSB['body'].head(5))

Case Folding Result :

22856  solusi nya bagaimana supaya bisa pakai email ...
20405  siapp terimakasih banyak admin 🙏
23076  baik, terima kasih atas informasinya
6747  salam sehat, jss support di sini. 🙏 😊 \nbila ad...
240    nik\n3471131004810001 \n\na.n. andri setiawan
Name: body, dtype: object

```

▼ Tokenizing

Menghapus karakter-karakter yang tidak penting dari text

```

import string
import re #regex library

# mengimport word_tokenize & FreqDist from NLTK
from nltk.tokenize import word_tokenize
from nltk.probability import FreqDist

# ----- Tokenizing -----

nltk.download('punkt')

```

```

def remove_tweet_special(text):
    # remove tab, new line, ans back slice
    text = str(text) # As the indicated in the error message, text must be string. You may try changing data type
    text = text.replace('\t', " ").replace('\n', " ").replace('\u', " ").replace('\ ', " ")
    # remove non ASCII (emoticon, chinese word, .etc)
    text = text.encode('ascii', 'replace').decode('ascii')
    # remove mention, link, hashtag
    text = ' '.join(re.sub("([@#][A-Za-z0-9+])|(\\w+:\\/\\S+)", " ", text).split())
    # remove incomplete URL
    return text.replace("http://", " ").replace("https://", " ")

dataSB['body'] = dataSB['body'].apply(remove_tweet_special)

#remove number
def remove_number(text):
    return re.sub(r"\d+", "", text)

dataSB['body'] = dataSB['body'].apply(remove_number)

#remove punctuation
def remove_punctuation(text):
    return text.translate(str.maketrans("", "", string.punctuation))

dataSB['body'] = dataSB['body'].apply(remove_punctuation)

#remove whitespace leading & trailing
def remove_whitespace_LT(text):
    return text.strip()

dataSB['body'] = dataSB['body'].apply(remove_whitespace_LT)

#remove multiple whitespace into single whitespace
def remove_whitespace_multiple(text):
    return re.sub('\s+', ' ', text)

dataSB['body'] = dataSB['body'].apply(remove_whitespace_multiple)

# remove single char

def remove_singl_char(text):
    return re.sub(r"\b[a-zA-Z]\b", "", text)

dataSB['body'] = dataSB['body'].apply(remove_singl_char)

# NLTK word tokenize
def word_tokenize_wrapper(text):
    return word_tokenize(text)

dataSB['body_tokens'] = dataSB['body'].apply(word_tokenize_wrapper)

print('Tokenizing Result : \n')
print(dataSB['body_tokens'].head())

[nltk_data] Downloading package punkt to /root/nltk_data...
[nltk_data] Unzipping tokenizers/punkt.zip.
Tokenizing Result :

22856  [solusi, nya, bagaimana, supaya, bisa, pakai,...
20405  [siapp, terimakasih, banyak, admin]
23076  [baik, terima, kasih, atas, informasinya]
6747  [salam, sehat, jss, support, di, sini, bila, a...
240    [nik, an, andri, setiawan]
Name: body_tokens, dtype: object

# NLTK calc frequency distribution
def freqDist_wrapper(text):
    return FreqDist(text)

dataSB['body_tokens_fdist'] = dataSB['body_tokens'].apply(freqDist_wrapper)

print('Frequency Tokens : \n')
print(dataSB['body_tokens_fdist'].head().apply(lambda x : x.most_common()))

Frequency Tokens :

22856  [(solusi, 1), (nya, 1), (bagaimana, 1), (supa...
20405  [(siapp, 1), (terimakasih, 1), (banyak, 1), (a...
23076  [(baik, 1), (terima, 1), (kasih, 1), (atas, 1)...

```

```

6747 [(jss, 5), (di, 2), (layanan, 2), (terkendala,...
240 [(nik, 1), (an, 1), (andri, 1), (setiawan, 1)]
Name: body_tokens_fdlist, dtype: object

```

▼ Stopwords

Membersihkan teks dari kata-kata yang termasuk ke dalam stopwords

```

nltk.download('stopwords')

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data] Unzipping corpora/stopwords.zip.
True

from nltk.corpus import stopwords

# ----- get stopword from NLTK stopword -----
# get stopword indonesia
list_stopwords = stopwords.words('indonesian')

# ----- menambahkan stopword secara manual -----
list_stopwords.extend([ " * " , " : " ,
    " ada " , " adalah " , " adanya " , " adapun " , " admin " , " aduan " , "
    " beri " , " berikan " , " berikut " , " berikutnya " , " berjumlah " , " berkali-
    " diberikan " , " diberikannya " , " dibuat " , " dibuatnya " , " didapat " ,
    " empat " , " enggak " , " enggan " , " entah " , " entahlah " , " erornya " , " erornya
    " kan " , " kapan " , " kapankah " , " kapanpun " , " karena " , " karenanya " , " kasih
    " manalagi " , " mangrimkan " , " mas " , " mas " , " masa " , " masalah " , " masalahnya "
    " menu " , " menuju " , " menunjuk " , " menunjuki " , " menunjukkan " , " menunjuknya " ,
    " pertanyaan " , " pertanyakan " , " pesan " , " pihak " , " pihaknya " , " proses " , " puku
    " saya " , " sekalian " , " sekaligus " , " sekali-kali " , " sekalipun " , " sekarang " , " seka
    " silahkan " , " sini " , " sinilah " , " smart " , " soal " , " soalnya " , " sore "
    " terlebih " , " saya " , " nan " , " sudah " , " yg " , " mohon " , " terlihat " , " termasuk " , " ternyata " , " ters
])

# convert list to dictionary
list_stopwords = set(list_stopwords)

#remove stopword pada list token
def stopwords_removal(words):
    return [word for word in words if word not in list_stopwords]

dataSB['body_tokens_WSW'] = dataSB['body_tokens'].apply(stopwords_removal)

print(dataSB['body_tokens_WSW'].head(10))

22856 [solusi, nya, bagaimana, pakai, email]
20405 [siapp, terimakasih, admin]
23076 [informasinya]
6747 [sehat, jss, support, layanan, jogja, smart, s...
240 [nik, an, andri, setiawan]
18572 [ok]
24476 [dipandu, jessica, terjawab, dibantu]
1009 [perkenalkan, amira, manajer, cabang, laborato...
17211 [malam, vaksin, tanggal, juli, kemarin, vaksin...
7333 [nuwun, koreksinya, mbak, iya, org]
Name: body_tokens_WSW, dtype: object

```

▼ Normalisasi

Mengganti kata-kata tertentu dengan kata lain yang lebih tepat

```
#import file xlsx untuk proses normalisasi
! gdown --id 1EQ_Vm4BzFuhNJbdLzRb-hvtvywRluxXb #id file di drive

/usr/local/lib/python3.8/dist-packages/gdown/cli.py:127: FutureWarning: Option `--id` was deprecated in version 4.3.1 &
warnings.warn(
Downloading...
From: https://drive.google.com/uc?id=1EQ\_Vm4BzFuhNJbdLzRb-hvtvywRluxXb
To: /content/normalisasi.xlsx
100% 8.54k/8.54k [00:00<00:00, 13.1MB/s]

normalizad_word = pd.read_excel('/content/normalisasi.xlsx') #lokasi file

normalizad_word_dict = {}

for index, row in normalizad_word.iterrows():
    if row[0] not in normalizad_word_dict:
        normalizad_word_dict[row[0]] = row[1]

def normalized_term(document):
    return [normalizad_word_dict[term] if term in normalizad_word_dict else term for term in document]

dataSB['body_normalized'] = dataSB['body_tokens_WSW'].apply(normalized_term)

dataSB['body_normalized'].head(10)

    22856          [solusi, nya, bagaimana, pakai, email]
    20405          [siapp, terimakasih, admin]
    23076          [informasinya]
    6747   [sehat, jss, support, layanan, jogja, smart, s...
    240          [nik, an, andri, setiawan]

    18572          [ok]
    24476   [dipandu, jessica, terjawab, dibantu]
    1009   [perkenalkan, amira, manajer, cabang, laborato...
    17211 [malam, vaksin, tanggal, juli, kemarin, vaksin...
    7333   [nuwun, koreksinya, mbak, iya, org]
    Name: body_normalized, dtype: object

# dataSB.isnull().sum()

# menghitung kata kunci dari chat

# import Sastrawi package
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import swifter

# create stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()

# stemmed
def stemmed_wrapper(term):
    return stemmer.stem(term)

term_dict = {}

for document in dataSB['body_normalized']:
    for term in document:
        if term not in term_dict:
            term_dict[term] = ''

print(len(term_dict))

313

for term in term_dict:
    term_dict[term] = stemmed_wrapper(term)
```

```

print(term, ":" , term_dict[term])

    suwun : suwun

# apply stemmed term to dataframe
def get_stemmed_term(document):
    return [term_dict[term] for term in document]

dataSB['body_tokens_stemmed'] = dataSB['body_normalized'].swifter.apply(get_stemmed_term)

print(dataSB['body_tokens_stemmed'])

Pandas Apply: 100%                                100/100 [00:00<00:00, 4026.91it/s]
22856          [solusi, nya, bagaimana, pakai, email]
20405          [siapp, terimakasih, admin]
23076          [informasi]
6747   [sehat, jss, support, layanan, jogja, smart, ser...
240          [nik, an, andri, setiawan]
...
19222   [sehat, jss, support, kendala, layanan, jogja, s...
12620          [aktivasi, nik]
24630          [udah, trimakasih]
5600   [bapakibu, ijin, urus, ubah, ktp, nggih, suwun]
1424          []
Name: body_tokens_stemmed, Length: 100, dtype: object

#stopwords #2 ada kata-kata lain yang belum masuk

from nltk.corpus import stopwords

# ----- get stopword from NLTK stopword -----
# get stopword indonesia
list_stopwords = stopwords.words('indonesian')

# ----- manually add stopword -----
list_stopwords.extend([ " * " , " : " ,
    " ada " , " adalah " , " adanya " , " adapun " , " admin " , " aduan " , " a
    " beri " , " berikan " , " berikut " , " berikutnya " , " berjumlah " , " berkali-
    " diberikan " , " diberikannya " , " dibuat " , " dibuatnya " , " didapat " ,
    " empat " , " enggak " , " enggaknya " , " entah " , " entahlah " , " erornya " , " erornya
    " kan " , " kapan " , " kapankah " , " kapanpun " , " karena " , " karenanya " , " kasih
    " manalagi " , " mangrimkan " , " mas " , " mas " , " masa " , " masalah " , " masalahnya "
    " menu " , " menuju " , " menunjuk " , " menunjuki " , " menunjukkan " , " menunjuknya " , "
    " pertanyaan " , " pertanyakan " , " pesan " , " pihak " , " pihaknya " , " proses " , " puku
    "sy" , " sekalian " , " sekaligus " , " sekali-kali " , " sekalipun " , " sekarang " , " sekarang
    " silahkan " , " sini " , " sinilah " , " smart " , " soal " , " soalnya " , " sore "
    " terlebih " , " terlihat " , " termasuk " , " ternyata " , " tersampaikan " , " tersebut "
])

# convert list to dictionary
list_stopwords = set(list_stopwords)

#remove stopword pada list token
def stopwords_removal(words):
    return [word for word in words if word not in list_stopwords]

dataSB['body_tokens_stemmed2'] = dataSB['body_tokens_stemmed'].apply(stopwords_removal)

print(dataSB['body_tokens_stemmed2'].head())

    22856          [solusi, nya, bagaimana, pakai, email]
    20405          [siapp, terimakasih, admin]
    23076          [informasi]
    6747   [sehat, jss, support, layanan, jogja, smart, ser...
    240          [nik, an, andri, setiawan]
Name: body_tokens_stemmed2, dtype: object

for i in range(len(dataSB)):
    a=dataSB.iloc[i][6]
    document.append(a)

document[0:5]

```

```

[['solusi', 'nya', 'bagaimana', 'pakai', 'email'],
 ['siapp', 'terimakasih', 'admin'],
 ['informasi'],
 ['sehat',
  'jss',
  'support',
  'layan',
  'jogja',
  'smart',
  'service',
  'kendala',
  'silah',
  'isi',
  'id',
  'jss',
  'afifa',
  'zakia',
  'ailana',
  'nama',
  'lengkap',
  'afifa',
  'zakia',
  'ailana',
  'nip',
  'nitb',
  'sbg',
  'pegawai',
  'nik',
  'sbg',
  'warga',
  'indonesia',
  'opd',
  'layan',
  'jss',
  'kendala',
  'akun',
  'jss',
  'eoffice',
  'ekinerja',
  'dll',
  'akun',
  'jss',

  'detail',
  'kendala',
  'alami',
  'ubah',
  'alamat',
  'lampir',
  'screenshot',
  'tangkap',
  'layar',
  'foto',
  'layar',
  'penuh',
  'tampil',
  'kendala',
  'eror',
  '...']]

doc_clean = dataSB['body_tokens_stemmed2']
doc_clean

22856      [solusi, nya, bagaimana, pakai, email]
20405      [siapp, terimakasih, admin]
23076      [informasi]
6747      [sehat, jss, support, layan, jogja, smart, ser...]
240        [nik, an, andri, setiawan]
...
19222     [sehat, jss, support, kendala, layan, jogja, s...]
12620      [aktivasi, nik]
24630      [udah, trimakasih]
5600      [bapakibu, ijin, urus, ubah, ktp, nggih, suwun]
1424      []
Name: body_tokens_stemmed2, Length: 100, dtype: object

```

▼ LDA model using gensim

Proses topic modeling dengan LDA gensim

```

import gensim
from gensim import corpora

```

```

dictionary = corpora.Dictionary(doc_clean)

dictionary.filter_extremes(no_below=2, no_above=0.9)
print(dictionary)

doc_term_matrix = [dictionary.doc2bow(doc) for doc in doc_clean]

    Dictionary<89 unique tokens: ['email', 'nya', 'solusi', 'terimakasih', 'informasi']...>

import warnings
warnings.filterwarnings('ignore')

import matplotlib.pyplot as plt
import re
import numpy as np
import pandas as pd
import gensim
import nltk

from nltk.stem.snowball import SnowballStemmer

from gensim.corpora import Dictionary, MmCorpus
from gensim.models import Ldamodel

from sklearn.cluster import KMeans
from sklearn.manifold import MDS
from sklearn.decomposition import PCA

import gensim.corpora as corpora
# Create Dictionary
id2word = corpora.Dictionary(doc_clean)
# Create Corpus
#texts = data_lemmatized
# Term Document Frequency
corpus = [id2word.doc2bow(text) for text in doc_clean]
# View
print(corpus[:1])

    [[(0, 1), (1, 1), (2, 1), (3, 1), (4, 1)]]

# Build LDA model
lda_model = gensim.models.LdaMulticore(corpus=corpus,
                                       id2word=id2word,
                                       num_topics=10,
                                       random_state=100,
                                       chunksize=100,
                                       passes=10,
                                       per_word_topics=True)

from pprint import pprint
# Print the Keyword in the 10 topics
pprint(lda_model.print_topics())
doc_lda = lda_model[corpus]

    [(0,
      '0.043*"ktp" + 0.023*"nama" + 0.023*"nik" + 0.023*"jss" + '
      '0.023*"terimakasih" + 0.023*"rubah" + 0.023*"informasi" + 0.023*"login" + '
      '0.023*"nikah" + 0.023*"data"'),
     (1,
      '0.093*"jss" + 0.056*"kendala" + 0.034*"layan" + 0.032*"layar" + '

```

```

'0.030*"akun" + 0.027*"jessica" + 0.023*"bantu" + 0.023*"pandu" + '
'0.021*"tangkap" + 0.020*"nama)'),
(2,
'0.028*"sdh" + 0.028*"error" + 0.028*"ceck" + 0.028*"rk" + 0.028*"gowongan" +
'+ 0.028*"freewifi" + 0.028*"balai" + 0.028*"jg" + 0.028*"file" + '
'0.028*"kompres)'),
(3,
'0.048*"jss" + 0.032*"akun" + 0.032*"nomor" + 0.032*"ganti" + 0.017*"sdh" + '
'0.017*"warga" + 0.017*"alamat" + 0.017*"sulit" + 0.017*"hilang" + '
'0.017*"mending)'),
(4,
'0.040*"jss" + 0.039*"ktp" + 0.039*"sewa" + 0.024*"jogja" + 0.016*"el" + '
'0.016*"aplikasi" + 0.016*"tsb" + 0.016*"vaksinasi" + 0.016*"nipitbnik" + '
'0.016*"pakbuk)'),
(5,
'0.063*"vaksin" + 0.032*"kemarin" + 0.031*"malam" + 0.017*"udah" + '
'0.017*"daftar" + 0.017*"konfir" + 0.017*"tpi" + 0.017*"terimakasihsalam" + '
'0.017*"sehat" + 0.017*"sertifnya)'),
(6,
'0.031*"email" + 0.031*"com" + 0.031*"solusi" + 0.017*"sistem" + '
'0.017*"kenal" + 0.017*"izin" + 0.016*"kak" + 0.016*"kait" + 0.016*"salah" + '
'0.016*"aip)'),
(7,
'0.034*"jss" + 0.032*"masuk" + 0.032*"nik" + 0.022*"vaksinasi" + '
'0.022*"perintah" + 0.022*"daftar" + 0.021*"eoffice" + 0.020*"sbg" + '
'0.015*"kendala" + 0.012*"warga)'),
(8,
'0.184*"lampir" + 0.177*"media" + 0.032*"jss" + 0.030*"warga" + 0.025*"akun" +
'+ 0.015*"ketua" + 0.015*"rw" + 0.015*"zakia" + 0.015*"afifa" + '
'0.015*"ailana)'),
(9,
'0.114*"daftar" + 0.049*"nik" + 0.025*"vaksin" + 0.025*"tulis" + 0.014*"jss" +
'+ 0.014*"nip" + 0.014*"ubah" + 0.014*"suwun" + 0.014*"urus" + '
'0.014*"bapakibu"')]

# Membuat objek untuk LDA menggunakan gensim
Lda = gensim.models.ldamodel.LdaModel

total_topics = 6 # jumlah topik yang akan di extract
number_words = 10 # jumlah kata per topik

# # config variables
# num_topics=4

# definisikan beberapa fungsi untuk kebutuhan pre-processing, pre-processing yang dilakukan adalah
# 1. lowercasing
# 2. stopword removal
# 3. stemming

stemmer = SnowballStemmer("english")
stopwords = nltk.corpus.stopwords.words('english')

def preprocess(text):

    # tokenizing and lowercasing
    tokens = [word.lower() for word in text.split()]
    filtered_tokens = []

    # buat yang bukan terdiri dari alfabet, dan merupakan stopword
    for token in tokens:
        if re.search('[a-zA-Z]', token) and (token not in stopwords):
            filtered_tokens.append(token)

    # lakukan stemming dengan snowball stemmer
    stems = [stemmer.stem(t) for t in filtered_tokens]
    return stems

# import gensim
# from gensim import corpora

# dictionary = corpora.Dictionary(doc_clean)

```



```

## Run the LDA !
# lda = LdaModel.LdaModel(corpus, num_topics=num_topics, id2word=dictionary, random_state=1, iterations=5000)

# Running and Trainign LDA model on the document term matrix.
lda_model = Lda(doc_term_matrix, num_topics=total_topics, id2word = dictionary, passes=50)

lda_model.show_topics(num_topics=total_topics, num_words=number_words)

[(0,
  '0.109*ny" + 0.059*ktp" + 0.059*alamat" + 0.059*ubah" + 0.008*solusi" + 0.008*email" + 0.008*udah" +
  0.008*rubah" + 0.008*sdh" + 0.008*aju"),
 (1,
  '0.136*daftar" + 0.083*vaksin" + 0.081*ktp" + 0.059*jss" + 0.049*warga" + 0.042*kemarin" + 0.042*yogyakarta"
  + 0.029*sdh" + 0.029*tulis" + 0.029*anak'),
 (2,
  '0.323*lampir" + 0.323*media" + 0.029*jogja" + 0.029*sulit" + 0.029*aju" + 0.016*surat" + 0.016*buka" +
  0.016*rencana" + 0.016*kenal" + 0.016*sistem'),
 (3,
  '0.149*nik" + 0.066*terimakasih" + 0.065*email" + 0.045*com" + 0.045*mbak" + 0.045*jss" + 0.045*solusi" +
  0.024*nomor" + 0.024*kak" + 0.024*ketua'),
 (4,
  '0.095*informasi" + 0.051*terimakasih" + 0.051*error" + 0.051*no" + 0.051*nik" + 0.051*nip" + 0.051*daftar" +
  0.007*udah" + 0.007*rubah" + 0.007*sdh'),
 (5,
  '0.110*jss" + 0.054*kendala" + 0.044*akun" + 0.037*layan" + 0.031*nama" + 0.027*layan" + 0.024*id" +
  0.024*lengkap" + 0.024*kirin" + 0.021*tangkap')]

# Word Count of Topic Keywords

from collections import Counter

topics = lda_model.show_topics(formatted=False)
data_flat = [w for w_list in doc_clean for w in w_list]
counter = Counter(data_flat)

out = []
for i, topic in topics:
    for word, weight in topic:
        out.append([word, i, weight, counter[word]])

df_imp_wcount = pd.DataFrame(out, columns=['word', 'topic_id', 'importance', 'word_count'])
print(df_imp_wcount)

   word  topic_id  importance  word_count
0     nya         0     0.109172          2
1     ktp         0     0.058854          8
2  alamat         0     0.058839          2
3     ubah         0     0.058839          2
4   solusi         0     0.008459          2
5   email         0     0.008439          4
6     udah         0     0.008406          2
7   rubah         0     0.008404          2
8     sdh         0     0.008403          3
9     aju         0     0.008403          2
10  daftar         1     0.136017         11
11  vaksin         1     0.082510          6
12     ktp         1     0.081289          8
13     jss         1     0.059061         39
14   warga         1     0.049409          6
15  kemarin         1     0.042370          3
16  yogyakarta     1     0.042369          3
17     sdh         1     0.029031          3
18   tulis         1     0.028990          2
19     anak         1     0.028990          2
20  lampir         2     0.323077         28
21   media         2     0.323036         24
22   jogja         2     0.028962          6
23   sulit         2     0.028876          2
24     aju         2     0.028699          2
25   surat         2     0.015598          2
26     buka         2     0.015595          2
27  rencana         2     0.015595          2

```

```

28      kenal      2      0.015591      2
29      sistem     2      0.015591      2
30      nik         3      0.149392     12
31      terimakasih 3      0.066265     4
32      email      3      0.065276     4
33      com        3      0.045383     3
34      mbak       3      0.045368     2
35      jss        3      0.045352     39
36      solusi     3      0.045351     2
37      nomor     3      0.024462     3
38      kak        3      0.024446     2
39      ketua     3      0.024439     3
40      informasi  4      0.094877     2
41      terimakasih 4      0.051196     4
42      error     4      0.051115     2
43      no        4      0.051100     3
44      nik        4      0.051088     12
45      nip        4      0.051079     4
46      daftar    4      0.051062     11
47      udah     4      0.007301     2
48      rubah     4      0.007300     2
49      sdh       4      0.007299     3
50      jss       5      0.110461     39
51      kendala   5      0.054246     16
52      akun      5      0.044179     13
53      layanan   5      0.037468     11
54      nama      5      0.030756     9
55      layar     5      0.027403     8
56      id        5      0.024046     7

```

```

#koneksi ke drive
from google.colab import drive
drive.mount('drive')

Drive already mounted at drive; to attempt to forcibly remount, call drive.mount("drive", force_remount=True).

#simpan ke google drive
df_imp_wcount.to_csv('df_imp_wcount.csv')
!cp df_imp_wcount.csv "drive/My Drive/File magang/LDA in colab 2/"

#Dominant topic and its percentage contribution in each topic
def format_topics_sentences(ldamodel=None, corpus=doc_term_matrix, texts=document):
    # Init output
    sent_topics_df = pd.DataFrame()

    # Get main topic in each document
    for i, row_list in enumerate(ldamodel[corpus]):
        row = row_list[0] if ldamodel.per_word_topics else row_list
        # print(row)
        row = sorted(row, key=lambda x: (x[1]), reverse=True)
        # Get the Dominant topic, Perc Contribution and Keywords for each document
        for j, (topic_num, prop_topic) in enumerate(row):
            if j == 0: # => dominant topic
                wp = ldamodel.show_topic(topic_num)
                topic_keywords = ", ".join([word for word, prop in wp])
                sent_topics_df = sent_topics_df.append(pd.Series([int(topic_num), round(prop_topic,4), topic_keywords]), ignore_index=True)
            else:
                break
    sent_topics_df.columns = ['Dominant_Topic', 'Perc_Contribution', 'Topic_Keywords']

    # Add original text to the end of the output
    contents = pd.Series(texts)
    sent_topics_df = pd.concat([sent_topics_df, contents], axis=1)
    return(sent_topics_df)

df_topic_sents_keywords = format_topics_sentences(ldamodel=lda_model, corpus=doc_term_matrix, texts=doc_clean)

# Format
df_dominant_topic = df_topic_sents_keywords.reset_index()
df_dominant_topic.columns = ['Document_No', 'Dominant_Topic', 'Topic_Perc_Contrib', 'Keywords', 'Text']
print(df_dominant_topic.head(10))

```

```

Document_No  Dominant_Topic  Topic_Perc_Contrib \
0            0            3.0            0.5415
1            1            3.0            0.5825
2            2            4.0            0.5833
3            3            5.0            0.9836
4            4            3.0            0.5828

```

```

5          5          0.0          0.1667
6          6          5.0          0.7917
7          7          2.0          0.9072
8          8          1.0          0.8958
9          9          3.0          0.5833

```

```

Keywords Text
0 nik, terimakasih, email, com, mbak, jss, solus... NaN
1 nik, terimakasih, email, com, mbak, jss, solus... NaN
2 informasi, terimakasih, error, no, nik, nip, d... NaN
3 jss, kendala, akun, layan, nama, layar, id, le... NaN
4 nik, terimakasih, email, com, mbak, jss, solus... NaN
5 nya, ktp, alamat, ubah, solusi, email, udah, r... NaN
6 jss, kendala, akun, layan, nama, layar, id, le... NaN
7 lampir, media, jogja, sulit, aju, surat, buka,... NaN
8 daftar, vaksin, ktp, jss, warga, kemarin, yogy... NaN
9 nik, terimakasih, email, com, mbak, jss, solus... NaN

```

```

#simpan ke google drive
df_dominant_topic.to_csv('df_dominant_topic.csv')
!cp df_dominant_topic.csv "drive/My Drive/File magang/LDA in colab 2/"

```

```

import pyLDAvis.gensim
import pickle
import pyLDAvis
# Visualize the topics
pyLDAvis.enable_notebook()

```

```

import os
LDAvis_data_filepath = os.path.join('ldavis_prepared_'+str(total_topics))

```

```

corpus = [dictionary.doc2bow(text) for text in doc_clean]

```

```

# if 1 == 1:
#     LDAvis_prepared = pyLDAvis.gensim.prepare(lda_model, corpus, dictionary)

```

```

# load the pre-prepared pyLDAvis data from disk
with open(LDAvis_data_filepath, 'rb') as f:
    LDAvis_prepared = pickle.load(f)

```

```

pyLDAvis.save_html(LDAvis_prepared, 'drive/My Drive/File magang/LDA in colab 2/ldavis_prepared_'+ str(total_topics) + '.html')

```

```

# proses scaling
LDAvis_prepared

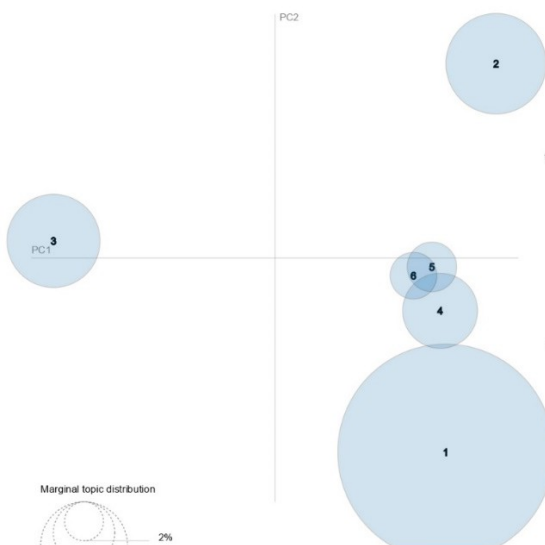
```

Selected Topic:

Slide to adjust relevance metric:⁽²⁾

$\lambda = 1$

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most E

