

## **Penerapan Metode *Text Mining* dengan *Chatbot Questions* *And Answer* pada PT PLN (Persero) Sumatera Selatan**

Abdur Rohim<sup>1</sup>, Muhammad Haviz Irfani<sup>1\*</sup>, Mustafa Ramadhan<sup>1</sup>, Ubaidillah  
Ubaidillah<sup>2</sup>

<sup>1</sup>Program Studi Teknik Informatika, Universitas Indo Global Mandiri

<sup>2</sup>Program Studi Ilmu Komputer, Universitas Sumatera Selatan

m.haviz@uigm.ac.id\*

---

### **Abstrak**

*Chatbot Questions and Answer* (Q&A) telah menjadi alat yang sangat penting bagi perusahaan untuk memberikan dukungan dan layanan pelanggan yang cepat dan efektif. Namun, tantangan utama dalam pengembangan *Chatbot Q&A* adalah memastikan bahwa chatbot dapat memahami bahasa alami pengguna dan memberikan jawaban yang tepat. Dalam penelitian ini, metode *Text Mining* dan *Natural Language Processing* (NLP) diterapkan pada *Chatbot Q&A* untuk PT PLN (Perusahaan Listrik Negara) Persero wilayah Sumatera Selatan dengan tujuan meningkatkan kemampuan *chatbot* dalam memahami pertanyaan pelanggan dan memberikan jawaban yang akurat. Aplikasi yang dirancang pada penelitian ini menggunakan metode pengembangan perangkat lunak yaitu model *Rational Unified Process* (RUP) dan diimplementasikan ke dalam sistem operasi android. Sedangkan untuk pengujiannya menggunakan perhitungan *Root Mean Square Error* (RMSE). Keakuratan pada pengujian ditunjukkan dengan hasil RMSE memiliki nilai kecil (mendekati nol). Dari hasil pengujian data sampel sejumlah 20 pertanyaan, nilai perhitungan RMSE sebesar 0.10, sehingga dapat disimpulkan tingkat akurasi prediksi pada proses *text mining* dalam aplikasi adalah sangat baik (RMSE mendekati 0). Hasil dari penelitian ini adalah alat bantu berupa aplikasi tanya-jawab informasi yang interaktif layaknya model diskusi dan dapat menggunakan bahasa sehari-hari.

**Kata kunci:** *Chatbot, Text Mining, Natural Language Processing, RMSE*

---

## 1. Pendahuluan

PLN merupakan sebuah Badan Usaha Milik Negara yang menyediakan jasa tenaga listrik untuk masyarakat dan telah memberikan kontribusi yang menangani masalah kepentingan listrik di Indonesia. PLN termasuk satu-satunya perusahaan penjual jasa listrik di Indonesia.

Informasi yang jelas dan mudah dipahami tentang layanan yang disediakan oleh PLN, seperti cara untuk mengajukan permintaan pemasangan listrik baru, pembayaran tagihan listrik, dan juga informasi tentang program energi bersih dan program *Corporate Social Responsibility* (CSR) yang dilakukan oleh PLN.

Sebagai BUMN unggul, PLN telah diakui sebagai perusahaan kelas dunia dengan perkembangan pesat dan terpercaya. Semua masyarakat Indonesia pun tentunya telah merasakan hasil kerja dari PLN, terutama melalui pasokan listrik yang mengalir setiap rumah-rumah tangga.

PLN harus aktif memberikan informasi kepada masyarakat tentang layanan dan kebijakan yang dilakukan oleh perusahaan, agar masyarakat dapat memanfaatkan layanan dengan lebih baik dan juga memahami peran penting PLN dalam penyediaan energi listrik bagi masyarakat. Informasi tentang layanan pelanggan sering dipublikasikan oleh PT PLN (Persero) Sumatera Selatan pada website dan media sosial. Rata-rata pengunjung atau *page views website pln.co.id* yaitu sebanyak 3 (tiga) juta setiap bulannya.

Teknologi Chatbot merupakan salah satu bentuk aplikasi *Natural Language Processing* (NLP) bidang ilmu Kecerdasan Buatan (*Artificial Intelligence*) yang mempelajari komunikasi antara manusia dengan komputer melalui bahasa alami. Model komputasi digunakan untuk memudahkan komunikasi antara manusia dengan komputer dalam hal pencarian informasi, sehingga dapat terjadi suatu interaksi antara keduanya dengan menggunakan bahasa alami. Harapannya

masyarakat dapat berinteraksi dan adanya tanya jawab langsung melalui teknologi ini untuk memperoleh informasi, khususnya informasi mengenai layanan dan kebijakan PT PLN.

## 2. Kajian Pustaka

### 2.1. Text Mining

Text mining adalah penggalian informasi dari teks oleh user menggunakan tools analisis. Secara umum text mining mengadopsi proses-proses di dalam data mining dan di dalam text mining juga menggunakan teknik data ([1], [2], [3]).

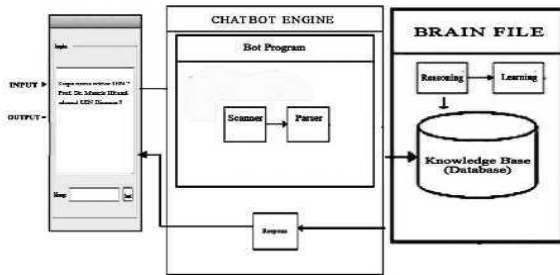
Dalam penerapan text mining, terdapat beberapa langkah yang perlu dilakukan antara lain :

- Tokenizing merupakan proses penguraian deskripsi yang semula berupa kalimat menjadi kata.
- Filtering adalah tahap mengambil kata penting dari hasil proses token. Bisa menggunakan algoritma *stop list* atau *word list*. Filtering dapat juga diartikan sebagai proses mengambil kata – kata penting dari hasil proses token atau penghapusan *stopwords* (kosa kata yang bukan merupakan ciri (kata unik) dari suatu dokumen).
- Stemming* adalah proses pemetaan dan penguraian berbagai bentuk (*variants*) dari suatu kata menjadi bentuk kata dasarnya (*stem*).
- Tagging* merupakan tahap untuk mencari bentuk awal/root dari tiap kata lampau atau hasil dari proses stemming.
- Analyzing* merupakan tahap penentuan seberapa jauh keterhubungan antar suatu kata atau term terhadap suatu dokumen atau kalimat dengan menghitung nilai/ bobot keterhubungan.

### 2.2. Arsitektur Chatbot

Penggunaan chatbot untuk memberikan rekomendasi produk atau menyelesaikan masalah pelanggan dapat meningkatkan kepuasan pelanggan dan memperkuat loyalitas merek[4]. Penting bagi perusahaan untuk memastikan bahwa chatbot mereka dirancang dengan tata bahasa yang ramah dan empati, sehingga pengalaman berinteraksi dengan

chatbot menjadi menyenangkan bagi pelanggan [5]. Chatbot terdiri dari dua komponen utama yakni *bot program* dan *brain file*.

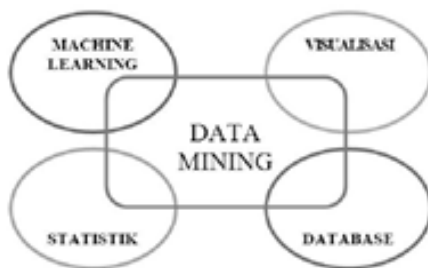


(Sumber: [6], [7])

Gambar 1. Arsitektur Chatbot

### 2.3 Data Mining

*Data Mining* adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam basis data. *Data mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar[5].



(Sumber: [2])

Gambar 2. Data Mining

*Knowledge Discovery in Database* merupakan proses terstruktur, yaitu sebagai berikut[3], [8]:

1. Pembersihan data (*data cleaning*), yaitu proses menghapus data pengganggu (*noise*) yang dikatakan tidak konsisten atau tidak diperlukan.
2. Integrasi data (*data integration*), yaitu menggabungkan berbagai sumber data.
3. Pemilihan data (*data selection*), yaitu memilih data yang dipilih sesuai kebutuhan analisis.

4. Transformasi data (*data transformation*), yaitu proses transformasi data ke dalam format untuk diproses dan siap ditambang.
5. Penggalan data (*data mining*), yaitu menerapkan metode kecerdasan untuk ekstraksi pola.
6. Evaluasi pola (*pattern evaluation*), yaitu mengidentifikasi pola-pola yang menarik yang merepresentasikan pengetahuan.
7. Penyajian pola (*knowledge presentation*), yaitu Teknik untuk memvisualisasikan pola pengetahuan ke pengguna.

### 2.4 Natural Language Processing

*Natural Language Processing* (Pemrograman Bahasa Alami) adalah pembuatan program yang memiliki kemampuan untuk memahami bahasa manusia. Pada prinsipnya bahasa alami adalah suatu bentuk representasi dari suatu pesan yang ingin dikomunikasikan antar manusia[9].

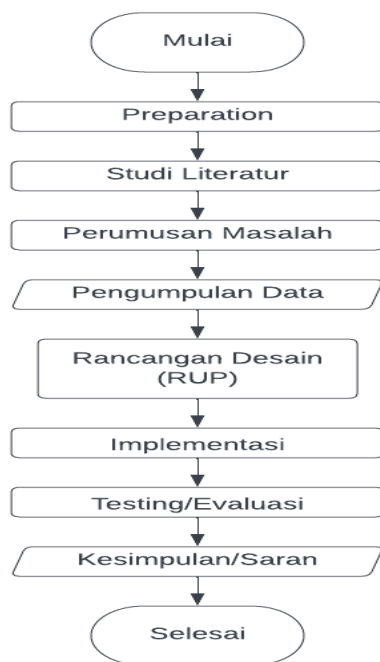
### 2.5 Android

Android adalah software besutan Google yang mencakup sistem operasi atau OS dan aplikasi kunci yang berbasis Linux pada suatu gadget seperti telephone cellular, smartphone dan komputer tablet.

Android studio adalah IDE (Integrated Development Environment) resmi untuk pengembangan aplikasi Android dan bersifat open source atau gratis. Peluncuran Android Studio ini diumumkan oleh Google pada 16 mei 2013 pada event Google I/O Conference untuk tahun 2013. Sejak saat itu, Android Studio menggantikan Eclipse sebagai IDE resmi untuk mengembangkan aplikasi Android.

### 3. Metode Penelitian

Tahapan penelitian ini dapat dilihat pada Gambar 1.



Gambar 3. Alur Penelitian

**Preparation (Persiapan):** merencanakan secara rinci, termasuk batasan penelitian.

**Studi Literatur:** memahami referensi atau literatur penelitian sebelumnya yang relevan dengan topik yang dikerjakan.

**Rumusan Masalah:** merumuskan pernyataan masalah secara jelas dan spesifik terkait dengan penerapan metode text mining pada chatbot questions and answer.

**Pengumpulan Data:** Data yang relevan dapat berupa percakapan pengguna dengan chatbot yang telah ada sebelumnya atau data yang telah diarsipkan.

**Rancangan Desain:** mencakup bagaimana penerapan metode text mining (termasuk text preprocessing) akan diintegrasikan ke dalam chatbot, merancang alur interaksi chatbot dan algoritma text mining.

**Implementasi:** Implementasi meliputi proses pembuatan kode, integrasi metode text mining, dan pengaturan sistem agar chatbot dapat beroperasi sesuai dengan yang diharapkan.

**Testing/Evaluasi:** melakukan pengujian dan evaluasi terhadap chatbot yang telah diimplementasikan.

**Kesimpulan:** menyimpulkan hasil

penelitian berdasarkan analisis data dan evaluasi yang telah dilakukan.

#### 4. Hasil dan Pembahasan

##### A. Proses Teks *Input*

Langkah awal dengan memasukkan teks pertanyaan yang sudah diberikan kode agar pemrosesan teks yang dimasukkan pengguna ke dalam bentuk kalimat yang mudah dipahami oleh mesin, teks tersebut dikumpulkan dari pertanyaan-pertanyaan pada chatbot, dan jawaban dari pertanyaan yang sudah dikategorikan juga sebelumnya (Gambar 2 dan Gambar 3).

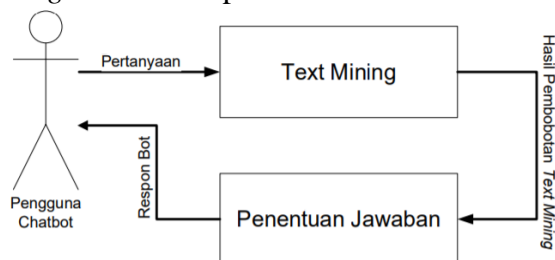
Kode Pertanyaan	Pertanyaan
P1	Apa saja layanan dan kebijakan dari PLN?
P2	Apakah ada biaya tambahan untuk pemasangan listrik di rumah baru?
P3	Apa saja syarat untuk mengajukan pemasangan listrik?
P4	Bagaimana cara mengganti nama pada rekening listrik PLN?
P5	Apakah ada promo atau diskon pembayaran tagihan listrik?
P6	Apakah bisa memindahkan sambungan listrik dari satu rumah ke rumah lain?
P7	Bagaimana cara menghubungi teknisi PLN jika terjadi gangguan listrik di rumah?
P8	Bagaimana cara mengajukan penggantian meteran listrik yang rusak?
P9	Bagaimana cara mendaftar sebagai pelanggan PLN?
P10	Apa yang harus dilakukan jika terjadi pemadaman listrik?
P11	Bagaimana cara mengajukan kompensasi jika terjadi kerusakan pada peralatan elektronik akibat gangguan listrik?
P12	Apakah bisa menggunakan solar panel untuk menghasilkan listrik?
P13	Bagaimana cara mengetahui jumlah daya listrik yang dibutuhkan untuk rumah?
P14	Apakah ada jenis tarif listrik yang tersedia di PLN?
P15	Apakah ada layanan perbaikan peralatan listrik yang rusak dari PLN?
P16	Bagaimana cara mengajukan permohonan penambahan daya listrik di rumah?
P17	Apakah bisa memilih jenis listrik yang digunakan, seperti listrik hijau atau listrik konvensional?
P18	Bagaimana cara mengetahui tagihan listrik yang harus dibayarkan?
P19	Bagaimana cara mengajukan surat pengaduan jika ada gangguan pelayanan dari PLN?
P20	Apa yang harus dilakukan jika terjadi kebocoran listrik di rumah?
P21	Apakah yang dimaksud dengan tarif adjustment?
P??	Apakah ada konsekuensi jika tidak membayar tagihan listrik?

Gambar 4. Pertanyaan dalam Chatbot

J18	Melihat tagihan yang dikirimkan ke alamat melalui pos atau kurir Melihat tagihan pada aplikasi PLN Mobile atau website resmi PLN Melihat tagihan pada mesin ATM atau internet banking Datang langsung ke kantor PLN terdekat dan menanyakan tagihan yang harus dibayarkan Pastikan untuk membayar tagihan listrik sebelum jatuh tempo agar terhindar dari denda atau pemutusan sementara pasokan listrik.
J19	Jika mengalami gangguan pelayanan dari PLN, dapat mengajukan surat pengaduan ke kantor PLN terdekat atau melalui website resmi PLN. Pastikan untuk mencantumkan informasi yang lengkap dan jelas mengenai gangguan yang terjadi, serta menyertakan bukti-bukti pendukung seperti nomor pelanggan, tanggal dan waktu terjadinya gangguan, serta nomor telepon yang dapat dihubungi. PLN akan menindaklanjuti pengaduan dan memberikan solusi yang terbaik dalam waktu yang sesingkat mungkin. juga dapat menghubungi call center PLN di nomor 123 untuk melaporkan gangguan pelayanan yang terjadi
J20	Jika terjadi kebocoran listrik di rumah, segera matikan semua peralatan listrik dan hubungi teknisi PLN di nomor 123 atau melalui aplikasi PLN Mobile. Jangan mencoba memperbaiki sendiri kebocoran listrik karena bisa sangat berbahaya. Teknisi PLN akan segera datang untuk memperbaiki kebocoran dan memastikan bahwa rumah aman untuk digunakan kembali.
J21	Tariff adjustment adalah mekanisme mengubah dan menetapkan turun naiknya besaran tarif listrik mengikuti perubahan besarnya faktor ekonomi mikro, agar tarif yang dikenakan kepada konsumen mendekati Biaya Pokok Penyediaan Listrik (BPP).
J22	Jika tidak membayar tagihan, maka pelanggan akan dikenakan denda keterlambatan pembayaran. Jika masih tidak dibayar dalam jangka waktu tertentu, maka PLN dapat melakukan pemutusan sementara pasokan listrik. Selain itu, jika tagihan

Gambar 5. Jawaban dalam Chatbot

Rancangan kebutuhan pengguna dengan metode NLP secara *usecase diagram* terlihat pada Gambar 4.



Gambar 6. Gambaran Umum Sistem Chatbot[1]

### B. Proses Tokenisasi Kalimat

Pemisahan kata perkata dari sebuah kalimat dalam kalimat pertanyaan dengan pemisah “-“ akan menghasilkan kata dasar, kata sambung, kata berimbuhan, karakter khusus yang akan mengawali terjadinya proses *filtering* (Gambar 5). Tokenisasi bisa berupa kata, frasa, simbol, atau unit lainnya yang memiliki makna. Seperti pertanyaan “Apakah bisa memindahkan sambungan listrik dari satu rumah ke rumah lain?” menjadi bentuk senarai ["Apakah", "bisa", "memindahkan", "sambungan", "listrik", "dari", "satu",

"rumah", "ke", "rumah", "lain", "?" ].

Kode Pertanyaan	Pertanyaan
P1	Apa saja layanan dan kebijakan dari PLN
P2	Apakah ada biaya tambahan untuk pemasangan listrik di rumah baru
P3	Apa saja syarat untuk mengajukan pemasangan listrik
P4	Bagaimana cara mengganti nama pada rekening listrik PLN
P5	Apakah ada promo atau diskon pembayaran tagihan listrik
P6	Apakah bisa memindahkan sambungan listrik dari satu rumah ke rumah lain
P7	Bagaimana cara menghubungi teknisi PLN jika terjadi gangguan listrik di rumah
P8	Bagaimana cara mengajukan penggantian meteran listrik yang rusak
P9	Bagaimana cara mendaftar sebagai pelanggan PLN
P10	Apa yang harus dilakukan jika terjadi pemadaman listrik
P11	Bagaimana cara mengajukan kompensasi jika terjadi kerusakan pada peralatan elektronik akibat gangguan listrik?
P12	Apakah bisa menggunakan solar panel untuk menghasilkan listrik
P13	Bagaimana cara mengetahui jumlah daya listrik yang dibutuhkan untuk rumah
P14	Apa saja jenis tarif listrik yang tersedia di PLN
P15	Apakah ada layanan perbaikan peralatan listrik yang rusak dari PLN
P16	Bagaimana cara mengajukan permohonan penambahan daya listrik di rumah

Gambar 7. Hasil Proses Tokenisasi

### C. Proses *Filtering* (Saringan) Kata

Dalam proses *filtering*, menghilangkan kata-kata yang tidak memberikan banyak informasi atau tidak relevan dalam konteks tertentu. Kata yang difilter seperti kata sambung, kata ganti, kata tanya, dan karakter khusus (Gambar 6).

Dalam contoh kalimat sebelumnya, melakukan *filtering* untuk menghapus kata-kata seperti "Apakah", "bisa", "dari", "satu", "ke", dan "?", karena kata-kata tersebut mungkin tidak memberikan informasi yang signifikan dalam menentukan maksud kalimat. Lain halnya untuk kata-kata seperti "memindahkan", "sambungan", "listrik", "rumah", dan "lain" yang relevan dan sebaiknya tetap dipertahankan. Hasil *filtering* dari token-token tersebut menjadi ["memindahkan", "sambungan", "listrik", "rumah", "lain"] yang mungkin memberikan informasi yang lebih berguna dalam menentukan makna dari kalimat sebelumnya.

Data Hasil Tokenizing	Hasil Filtering
Apa - saja - syarat - untuk - mengajukan - pemasangan - listrik	syarat - mengajukan - pemasangan - listrik
Bagaimana - cara - mengganti - nama - pada - rekening - listrik - PLN	cara - mengganti - nama - rekening - listrik - PLN
Apakah - ada - promo - atau - diskon - pembavaran - tagihan - listrik	promo - diskon - pembayaran - tagihan - listrik
Apakah - bisa - memindahkan - sambungan - listrik - dari - satu - rumah - ke - rumah - lain	memindahkan - sambungan - listrik - satu - rumah - ke - rumah - lain
Bagaimana - cara - menghubungi - teknisi - PLN - jika - terjadi - gangguan - listrik - di - rumah	cara - menghubungi - teknisi - PLN - jika - terjadi - gangguan - listrik - di - rumah
Bagaimana - cara - mengajukan - penggantian - meteran - listrik - yang - rusak	cara - mengajukan - penggantian - meteran - listrik - rusak
Bagaimana - cara - mendaftar - sebagai - pelanggan - PLN	cara - mendaftar - sebagai - pelanggan - PLN
Apa - yang - harus - dilakukan - jika - terjadi - pemadaman - listrik	harus - dilakukan -Jika - terjadi - pemadaman - listrik
Bagaimana - cara - mengajukan - kompensasi - jika - terjadi - kerusakan - pada - peralatan - elektronik - akibat - gangguan - listrik	cara - mengajukan - kompensasi - jika - terjadi - kerusakan - peralatan - elektronik - akibat - gangguan - listrik
Apakah - bisa - menggunakan - solar - panel - untuk - menghasilkan - listrik	menggunakan - solar - panel - menghasilkan - listrik
Bagaimana - cara - mengetahui - jumlah - daya - listrik - yang - dibutuhkan - untuk - rumah	cara - mengetahui - jumlah - daya - listrik - dibutuhkan - rumah
Apa - saja - jenis - tarif - listrik - yang - tersedia - di - PLN	jenis - tarif - listrik - tersedia - di - PLN
Apakah - ada - layanan - perbaikan - peralatan - listrik - yang - rusak - dari - PLN	layanan - perbaikan - peralatan - listrik - rusak - PLN
Bagaimana - cara - mengajukan - permohonan - penambahan - daya - listrik - di - rumah	cara - mengajukan - permohonan - penambahan - daya - listrik - di - rumah
Apakah - bisa - memilih - jenis - listrik - yang - digunakan - seperti - listrik - hijau -	memilih - jenis - listrik - digunakan - seperti - listrik -

Gambar 8. Hasil Proses *Filtering*

#### D. Proses *Stemming* Kata

Proses mengubah kata-kata menjadi bentuk dasarnya atau "akar kata", sehingga kata-kata dengan akar yang sama dianggap identik. Ini membantu dalam mengurangi variasi kata yang sama dalam teks. Dalam contoh yang sama, menggunakan algoritma stemming untuk mengubah kata-kata ke dalam bentuk dasarnya (Gambar 7). Berikut adalah hasil stemming untuk kata-kata yang telah difilter sebelumnya:

"memindahkan" menjadi "mindah"

"sambungan" menjadi "sambung"

"listrik" tetap "listrik"

"rumah" tetap "rumah"

"lain" tetap "lain"

Hasil akhir setelah proses stemming adalah:

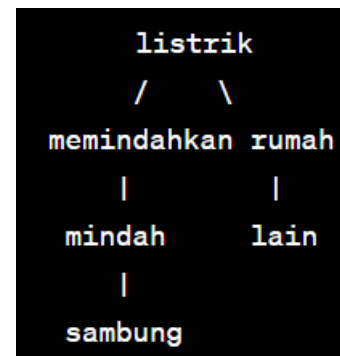
["mindah", "sambung", "listrik", "rumah", "lain"]

Hasil Filtering	Hasil Stemming
listrik - di - rumah - baru	listrik -rumah - baru
syarat - mengajukan - pemasangan - listrik	syarat - ajukan - pemasangan - listrik
cara - mengganti - nama - rekening - listrik - PLN	cara - ganti - nama - rekening - listrik - PLN
promo - diskon - pembayaran - tagihan - listrik	promo - diskon - pembayaran - tagihan - listrik
memindahkan - sambungan - listrik - satu - rumah - ke - rumah - lain	mindah - sambungan - listrik - satu - rumah - rumah - lain
cara - menghubungi - teknisi - PLN - jika - terjadi - gangguan - listrik - di - rumah	cara - hubungi - teknisi - PLN - jika - terjadi - gangguan - listrik - rumah
cara - mengajukan - penggantian - meteran - listrik - rusak	cara - ajukan - ganti - meteran - listrik - rusak
cara - mendaftar - sebagai - pelanggan - PLN	cara - daftar - sebagai - pelanggan - PLN
harus - dilakukan -Jika - terjadi - pemadaman - listrik	harus - dilakukan -Jika - terjadi - pemadaman - listrik
cara - mengajukan - kompensasi - jika - terjadi - kerusakan - peralatan - elektronik - akibat - gangguan - listrik	cara - ajukan - kompensasi - jika - terjadi - kerusakan - peralatan - elektronik - akibat - gangguan - listrik
menggunakan - solar - panel - menghasilkan - listrik	gunakan - solar - panel - menghasilkan - listrik
cara - mengetahui - jumlah - daya -	cara - tahu - jumlah - daya - listrik -

Gambar 9. Hasil Proses *Stemming*

#### E. Pola yang dihasilkan

Berikut ini graf hubungan kata-kata setelah proses stemming sebelumnya.



Gambar 8. Contoh graf setelah proses stemming

Graf tersebut hanya memvisualkan, juga dapat memberikan interpretasi bagi teks kalimat pertanyaan lainnya yang sudah terdaftar (Gambar 9).

Kode Pertanyaan	Pola
P1	Layanan[1] - kebijakan[1] - PLN[1]
P2	biaya[1] - tambahan[1] - pemasangan[1] - listrik[1] - rumah[1] - baru[1]
P3	syarat[1] - ajukan[1] - pemasangan[1] - listrik[1]
P4	cara[1] - ganti[1] - nama[1] - rekening[1] - listrik[1] - PLN[1]
P5	promo[1] - diskon[1] - pembayaran[1] - tagihan[1] - listrik[1]
P6	pindah[1] - sambungan[1] - listrik[1] - satu[1] - rumah[1] - rumah[1] - lain[1]
P7	cara[1] - hubungi[1] - teknisi[1] - PLN[1] - jika[1] - terjadi[1] - gangguan[1] - listrik[1] - rumah[1]
P8	cara[1] - ajukan[1] - ganti[1] - meteran[1] - listrik[1] - rusak[1]
P9	cara[1] - daftar[1] - sebagai[1] - pelanggan[1] - PLN[1]
P10	harus[1] - dilakukan[1] - jika[1] - terjadi[1] - pemadaman[1] - listrik[1]
P11	cara[1] - ajukan[1] - kompensasi[1] - jika[1] - terjadi[1] - kerusakan[1] - peralatan[1] - elektronik[1] - akibat[1] - gangguan[1] - listrik[1]
P12	gunakan[1] - solar[1] - panel[1] - menghasilkan[1] - listrik[1]
P13	cara[1] - tahu[1] - jumlah[1] - daya[1] - listrik[1] - kebutuhan[1] - rumah[1]
P14	jenis[1] - tarif[1] - listrik[1] - tersedia[1] - PLN[1]
P15	layanan[1] - perbaikan[1] - peralatan[1] - listrik[1] - rusak[1] - PLN[1]
P16	cara[1] - ajukan[1] - permohonan[1] - penambahan[1] - daya[1] - listrik[1] - rumah[1]
P17	pilih[1] - jenis[1] - listrik[1] - digunakan[1] - seperti - listrik - hijau[1] - atau[1] - listrik[1] - konvensional[1]
P18	cara[1] - tahu[1] - tagihan[1] - listrik[1] - harus[1] - listrik[1]

Gambar 10. Hasil proses *Stemming* dari teks pertanyaan

#### F. Perhitungan TF-IDF

Untuk menghitung TF-IDF dari hasil stemming sebelumnya, Hitung Term Frequency (TF): menghitung seberapa sering setiap kata muncul dalam Gambar 9 tersebut.

$$TF(t, d) = \frac{\text{jumlah kemunculan kata } t \text{ dalam dokumen } d}{\text{total kata dalam dokumen } d}$$

Di mana t adalah kata yang ingin dihitung TF-nya, d adalah dokumen, dan TF(t,d) adalah frekuensi kata t dalam dokumen d.

Sebagai contoh sebelumnya memiliki kata-kata: ["mindah", "sambung", "listrik", "rumah", "lain"]. Misalkan memiliki dokumen tunggal yang berisi kata-kata tersebut. Karena hanya memiliki satu dokumen, IDF untuk setiap kata akan sama.

$$IDF(t, D) = \log \left( \frac{\text{total dokumen dalam koleksi } D}{1 + \text{jumlah dokumen yang mengandung kata } t} \right)$$

Di mana D adalah koleksi dokumen, t adalah kata yang ingin dihitung IDF-nya, dan IDF(t,D) adalah nilai IDF untuk kata t.

Misalnya, jika memiliki total 100 dokumen dalam koleksi, dan sebuah kata muncul dalam 10 dokumen, maka IDF untuk kata tersebut Hitung TF-IDF: Setelah

mendapatkan nilai TF dan IDF untuk setiap kata, mendapatkan nilai TF-IDF[5].

$$TF-IDF(t, d, D) = TF(t, d) \times IDF(t, D)$$

Misalnya, jika TF untuk kata "mindah" adalah 0.2 dan IDF untuk kata "mindah" adalah 1.5, maka nilai TF-IDF untuk kata "mindah" dalam dokumen tersebut adalah  $0.2 * 1.5 = 0.3$ .

No	Token	IDF
1	Ajukan	0,57
2	Akibat	1,48
3	Baru	1,48
4	Cara	0,57
5	Daftar	1,48
6	Daya	1,18
7	Dilakukan	1,18
8	Diskon	1,48
9	Distribusi	1,48
10	Elektronik	1,48
11	Gangguan	1,00
12	Ganti	1,18
13	Gardu	1,48
14	Harus	1,00
15	Hubungi	1,48
16	Jenis	1,18
17	Jika	0,78
18	Jumlah	1,48
19	Kebijakan	1,48
20	Kebutuhan	1,48
21	Kerusakan	1,48

No	Token	IDF
32	Panel	1,48
33	Peduli	1,48
34	Pelanggan	1,48
35	Pemadaman	1,48
36	Pembayaran	1,00
37	Pembayaran	1,48
38	Penambahan	1,48
39	Pengaduan	1,48
40	Peralatan	1,00
41	Perbaikan	1,18
42	Permohonan	0,88
43	Pindah	1,48
44	PLN	0,44
45	Promo	1,48
46	Rekening	1,18
47	Rumah	0,70
48	Rusak	1,00
49	Sambungan	1,48
50	Satu	1,48
51	Sebagai	1,48
52	Solar	1,48

Gambar 11. Hasil perhitungan IDF

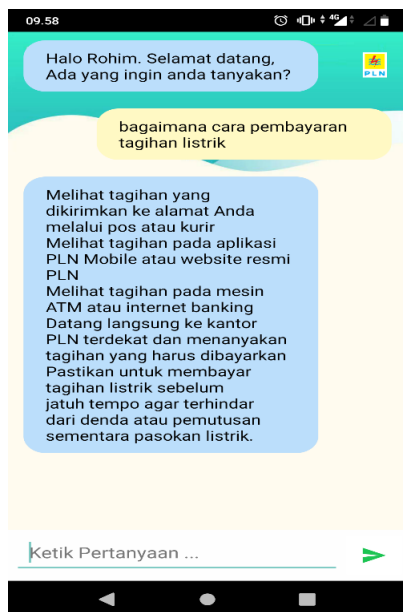
Kode Pertanyaan	Layanan	Kebijakan	PLN	Σ	IDF	W	Kode Jawaban
P1	1	1	1	3	1,00	3	J1
P2	0	0	0	0	0	0	J2
P3	0	0	0	0	0	0	J3
P4	0	0	1	1	1,48	1,48	J4
P5	0	0	0	0	0	0	J5
P6	0	0	0	0	0	0	J6
P7	0	0	1	1	1,48	1,48	J7
P8	0	0	0	0	0	0	J8
P9	0	0	1	1	1,48	1,48	J9
P10	0	0	0	0	0	0	J10
P11	0	0	0	0	0	0	J11
P12	0	0	0	0	0	0	J12
P13	0	0	0	0	0	0	J13
P14	0	0	1	1	0	0	J14
P15	0	0	1	1	1,48	1,48	J15
P16	0	0	0	0	0	0	J16
P17	0	0	0	0	0	0	J17
P18	0	0	0	0	0	0	J18
P19	0	0	1	1	1,48	1,48	J19
P20	0	0	0	0	0	0	J20
P21	0	0	0	0	0	0	J21
P22	0	0	0	0	0	0	J22
P23	0	0	1	1	1,48	1,48	J23
P24	0	0	1	1	1,48	1,48	J24

Gambar 12. Hasil perhitungan TF-IDF

#### G. Visual Aplikasi Chatbot



Gambar 13. Tampilan Halaman Menu Utama



Gambar 14. Tampilan Antar Muka Halaman Chatbot

Setelah dilakukan pengujian pada 20 dari 30 pertanyaan terdapat 2 keluaran yang berbeda, dalam menentukan perhitungan ini menggunakan akar kesalahan kuadrat rata-rata (*Root Mean Square Error*) sebagai berikut :

$$\begin{aligned} \text{RMSE} &= \sqrt{\frac{(2)^2}{20(20-1)}} \\ &= \sqrt{4/380} \\ &= 0.10 \end{aligned}$$

Berdasarkan hasil perhitungan RMSE di atas, nilai yang didapat adalah 0.10 dan dapat disimpulkan bahwa tingkat akurasi hasil prediksi pada proses text mining secara umum sudah baik hal itu dapat dilihat dari tingkat kesalahan yang relatif kecil.

## 5. Kesimpulan

Pendekatan Natural Language Processing dalam membangun aplikasi percakapan tersebut terjadi layaknya dilakukan antara manusia dengan manusia dan dalam penerapan metode *Text Mining* sebagai penalaran kalimat lebih memudahkan dalam pencarian *keyword* (*pattern*). Dari hasil pengujian data sampel sejumlah 20 pertanyaan nilai perhitungan RMSE sebesar 0.10, sehingga dapat disimpulkan tingkat akurasi prediksi pada proses text mining baik (RMSE mendekati 0).

## Referensi

- [1] M. D. Illescas-Manzano, N. V. López, N. A. González, and C. C. Rodríguez, "Implementation of chatbot in online commerce, and open innovation," *Journal of Open Innovation: Technology, Market, and Complexity*, vol. 7, no. 2, Jun. 2021, doi: 10.3390/joitmc7020125.
- [2] A. Gede, S. Pradnyana, M. Kom, K.



- Kom, S. Agustini, and M. S. Si, “Konsep Dasar Data Mining.”
- [3] D. Oleh, “MODUL DATA MINING TEXT MINING PERTEMUAN 12 (ONLINE).”
- [4] A. R. Dwi, F. Imamah, Y. S. Mei Andre, F. Ilmu Komputer, U. Mercu Buana Jalan Menteng Raya No, and K. Sirih, “APLIKASI CHATBOT (MILKI BOT) YANG TERINTEGRASI DENGAN WEB CMS UNTUK CUSTOMER SERVICE PADA UKM MINSU,” *Jurnal Cendikia*, vol. 16, no. 1, pp. 216–9436, 2018, [Online]. Available: <http://www.infodanpengertian.com/pengertian->
- [5] D. C. Algoritma, T. Setiadi, and L. Zahrotun, “Text Mining Untuk Mengklasifikasi Judul Skripsi Menggunakan Metode TF IDF,” 2018. [Online]. Available: <http://journal.uad.ac.id/index.php/JSTIF>
- [6] E. N. S. C. P and I. Afrianto, “Rancang Bangun Aplikasi Chatbot Informasi Objek Wisata Kota Bandung Dengan Pendekatan Natural Language Processing,” *Komputa : Jurnal Ilmiah Komputer dan Informatika*, vol. 4, no. 1, pp. 49–54, 2015, doi: 10.34010/komputa.v4i1.2410.
- [7] E. Nila and I. Afrianto, “RANCANG BANGUN APLIKASI CHATBOT INFORMASI OBJEK WISATA KOTA BANDUNG DENGAN PENDEKATAN NATURAL LANGUAGE PROCESSING,” *Jurnal Ilmiah Komputer dan Informatika (KOMPUTA)*, vol. 49, no. 1, 2015, [Online]. Available: [www.bandungtourism.com](http://www.bandungtourism.com).
- [8] A. Firdaus and W. I. Firdaus, “Text Mining Dan Pola Algoritma Dalam Penyelesaian Masalah Informasi : (Sebuah Ulasan),” 2021.
- [9] M. Astiningrum, P. Yoga Saputra, and M. Shoburu Rohmah, “IMPLEMENTASI NLP DENGAN KONVERSI KATA PADA SISTEM CHATBOT KONSULTASI LAKTASI,” *Jurnal Informatika Polinema*, vol. 5, no. 1, pp. 46–52, 2018.