

# BAB 1

## USULAN GAGASAN

### 1.1 Deskripsi Umum Masalah

Berkembangnya era digital yang semakin canggih, jumlah data yang dihasilkan oleh perusahaan di berbagai *platform online* seperti media sosial, situs ulasan, dan forum diskusi semakin meningkat secara eksponensial. Data ini mencakup berbagai pendapat dan opini perasaan pengguna *platform online* terhadap kualitas layanan perusahaan yang diberikan. PT PLN (Persero) adalah salah satu perusahaan yang menyediakan layanan kepada pelanggan menggunakan aplikasi yaitu *PLN Mobile*. Aplikasi tersebut bebas diunduh di *Google Play Store* untuk pengguna *Android*. Aplikasi *PLN Mobile* mengakibatkan adanya tanggapan dari pengguna aplikasi terhadap layanan yang diberikan oleh aplikasi, tanggapan tersebut dapat ditulis bebas pada bagian komentar di *platform Play Store*. Pada ulasan tersebut terdapat *Big Data* karena ulasan meningkat secara cepat dan bervariasi, sehingga PT PLN (Persero) kesulitan saat melakukan pengolahan data secara konvensional.

Analisis sentimen dapat membantu perusahaan untuk mengolah data dari ulasan pelanggan dalam merespons produk mereka, memonitor reputasi merek, dan merencanakan strategi pemasaran. Analisis sentimen adalah sebuah metode untuk mengumpulkan, mengelompokkan, mengelola, dan menganalisis opini dan sentimen dalam teks secara komputasional [1]. *Machine Learning* dipakai untuk menyusun model prediktif yang mampu mengelompokkan sentimen menjadi positif, negatif, atau netral berdasarkan data historis atau data *Training*. Kemudian, model-model ini digunakan untuk menganalisis dan memproses data baru yang diperoleh dari sumber seperti media sosial, ulasan produk, atau survei pelanggan [2]. Oleh karena itu, penelitian ini berfokus untuk membuat sistem klasifikasi teks otomatis pada ulasan aplikasi *PLN Mobile* berdasarkan sentimen dan topiknya menggunakan *Machine Learning*. Hasil dari sistem tersebut berupa data perbandingan jumlah sentimen tiap topik yang diharapkan dapat menjadi *insight* bisnis untuk PT PLN (Persero) dalam meningkatkan pelayanan kepada masyarakat.

## 1.2 Analisis Masalah

Berdasarkan latar belakang yang telah dituliskan di atas, maka beberapa aspek Analisis masalah pada penelitian ini adalah sebagai berikut:

### 1.2.1 Aspek Teknis

Memanfaatkan teknik *Machine Learning* untuk menganalisa rating dan review dapat membantu brand mengoptimalkan pemasaran dan layanannya secara lebih akurat dan efektif. Proses yang memudahkan analisis ini dilakukan dengan menampilkan sentimen dari review pengguna *platform* yang dapat membantu brand menganalisis dengan cepat tanggapan pengguna *platform* pada produk yang telah dipasarkan, sehingga brand dapat membuat keputusan yang lebih informatif dan strategis.

### 1.2.2 Aspek Ekonomi

Dalam permasalahan ekonomi sistem analisis sentimen dapat memberikan manfaat ekonomi yang signifikan bagi brand atau perusahaan, terutama dalam mengurangi biaya operasional, meningkatkan efisiensi dan efektivitas, serta menciptakan peluang pemasaran dan pengambilan keputusan yang lebih baik. Hal ini dikarenakan proses pengklasifikasian data teks menggunakan *Machine Learning* adalah proses yang cepat, akurat dan skalabel, sehingga dapat mengurangi biaya operasional dan meningkatkan efisiensi dan efektivitas dalam mengolah dan menganalisis sumber data yang besar.

### 1.2.3 Aspek Keberlanjutan

Dengan adanya analisis sentimen dengan *Machine Learning* diharapkan dapat menjadi solusi yang efektif dan berkelanjutan untuk mempermudah masyarakat khususnya pengguna teknologi digital dalam perusahaan maupun individu untuk mengklasifikasikan komentar atau ulasan yang bersifat positif dan negatif.

### 1.3 Analisis Solusi yang Ada

Penelitian tentang analisis sentimen dengan *Machine Learning* sebagai sistem klasifikasi data berupa teks, yang dapat memprediksi sentimen positif, negatif, dan netral sudah dilakukan oleh peneliti terdahulu. Penelitian mengenai metode *topic modeling* yang dapat memprediksi suatu topik pada kalimat juga sudah dilakukan oleh peneliti terdahulu, dari penelitian terdahulu data yang digunakan sangat bervariasi seperti ulasan pada suatu aplikasi pada *platform Play Store, twitter, online shop*, dan masih banyak lagi. Karena para peneliti terdahulu sudah membuktikan bahwa metode *Machine Learning* dan *topic modeling* dapat dijadikan solusi untuk melakukan klasifikasi pada teks, maka kami memilih metode tersebut pada penelitian ini. Beberapa contoh penelitian tentang sentimen analisis menggunakan metode *Machine Learning* dan *topic modeling* akan dipaparkan di bawah ini:

#### 1.3.1 Metode Analisis Sentimen Menggunakan *Machine Learning*.

Metode analisis sentimen dengan *Machine Learning* dilakukan dengan menggunakan sekumpulan algoritma yang digunakan untuk melatih set data historis dan membuat prediksi masa depan. Alasan pemilihan metode analisis sentimen dengan *Machine Learning* karena cenderung lebih akurat dibandingkan *sentiment dictionary*, tetapi bergantung pada kualitas dari corpus yang dilabel dengan polaritas [3]. Berikut adalah contoh peneliti terdahulu mengenai sentimen analisis menggunakan *Machine Learning*.

##### 1.3.1.1 Analisis sentimen pada Media Sosial *Twitter* Terhadap Kebijakan Pemberlakuan.

Pembatasan Kegiatan Masyarakat Berbasis *Deep Learning* oleh Mohammad Farid Nufal dan Selvia Ferdiana Kusuma [4].

Berdasarkan penelitian tersebut telah berhasil mengklasifikasikan 37759 tweet sesuai dengan sentimennya dengan akurasi 87%. Pengambilan data pada penelitian tersebut menggunakan metode *crawling*. Algoritma yang digunakan adalah *Long Short-Term Memory (LSTM)*.

##### 1.3.1.2 Penerapan metode Naive Bayes pada analisis sentimen Aplikasi McDonalds di Google *Play Store* oleh Rohannisa Fathwa Daud, Bambang Irawan, dan Agus Bahtiar [5].

Penelitian tersebut menggunakan 1.000 ulasan aplikasi *McDonalds* di *Play Store* dengan menggunakan algoritma *Naïve Bayes*. Hasil dari penelitian tersebut adalah model mendapat nilai akurasi sebesar 88% dengan nilai *Precision 90%*, *Recall 95%* dan nilai *F1-Score 93%*.

1.3.1.3 Review Analisis Sentimen Pelanggan Aplikasi Ruang Guru Menggunakan Metode BERT (*BiDirectional Encoder Representations from Transformers*) oleh Raden Mas Riky, Wahyu Panca Kusuma Atmaja, dan Wiyli Yutanti [6].

Pada penelitian tersebut menggunakan 5.437 ulasan aplikasi *Ruang Guru* dengan menggunakan model *Transformers Indo-BERT (Bidirectional Encoder Representations from Transformers)*. Hasil dari penelitian tersebut model mendapatkan nilai akurasi sebesar 99% dengan nilai *Precision 64,13%*, *Recall 60,51*, dan *F1-Score 98,1%*.

1.3.1.4 Analisis Sentimen Berbasis Aspek Pada Ulasan Aplikasi MYBLUBIRD Dengan Implementasi N-Gram Dan Algoritma *Logistic Regression* [7].

Penelitian ini menggunakan 873 ulasan berbahasa Inggris dari total 1.564 ulasan aplikasi MyBluebird. Proses *topic modeling* menggunakan LDA (*Latent Dirichlet Allocation*) menghasilkan tiga topik utama: taxi user experience, function app, dan user interface dengan *Coherence Score 0,469*. Klasifikasi sentimen dilakukan menggunakan *Logistic Regression* dengan fitur N-Gram, di mana model unigram menghasilkan performa terbaik (*Accuracy 88,8%*, *Precision 97,8%*, *Recall 87,7%*, *F1-Score 92,5%*). Analisis aspek menunjukkan mayoritas ulasan positif, dengan masukan terkait layanan, fitur aplikasi, dan desain antarmuka.

1.3.2 Metode *Topic Modeling*.

Pemodelan topik atau *Topic modeling* adalah satu metode *unsupervised Machine Learning* untuk mengorganisasi teks. Algoritma pemodelan topik menyediakan teknik untuk mengelompokkan tema sebagai topik. Pemodelan topik termasuk dalam *soft/fuzzy clustering* yang mana setiap objek dapat dimiliki lebih dari satu cluster [9]. Karena metode *topic modeling* mempunyai sistem yang dapat mengelompokkan topik pada data, maka kami memilih metode *topic modeling* untuk ditambahkan menjadi fitur tambahan pada penelitian ini. Berikut adalah beberapa peneliti terdahulu mengenai *Topic modeling* dengan model algoritma yang berbeda:

1.3.2.1 *Topic modeling* Pada Abstrak Skripsi Menggunakan Metode *Latent Semantic Analysis* oleh Rifki Hakim, Khalid, dan Dwi Rolliawati [10].

Pada penelitian tersebut menggunakan 720 abstrak skripsi untuk diklasifikasikan topiknya. Penelitian ini menggunakan model *LSA (Latent Semantic Analysis)* dengan hasil terdapat 2 topik yang dibedakan yaitu linguistik dan literatur, selanjutnya dari tiap topik tersebut terdapat topik didalamnya yang berjumlah 7 topik yang disajikan dalam tren pertahun dari tahun 2014 sampai dengan 2019.

### 1.3.2.2 Aplikasi Topiv Modeling Pada Pemberitaan Portal Berita *Online* Selama Masa PSBB Pertama Oleh Wahyudin [11]

Penelitian tersebut menggunakan 59.279 berita *online* untuk dijadikan dataset. Model yang digunakan adalah *LDA* dengan hasil penelitian mendapatkan 20 topik yang dikelompokkan menjadi 4 topik besar dengan melakukan topik kogerensi dan pravalensi dari tiap topik tersebut.

## 1.4 Tujuan Tugas Akhir

Tujuan dari Tugas Akhir ini adalah untuk merancang dan mengimplementasikan sistem analisis sentimen berbasis *Machine Learning* dan *topic modeling* guna mengklasifikasikan opini pengguna terhadap aplikasi PLN *Mobile* pada *platform Play Store*. Adapun tujuan khusus dari penelitian ini antara lain:

1. Mengembangkan sistem klasifikasi sentimen otomatis terhadap ulasan pengguna aplikasi PLN *Mobile* menggunakan algoritma *Machine Learning* dan *Deep Learning* seperti *Transformer IndoBERT*, *Naïve Bayes*, *LSTM*, dan *Logistic Regression*.
2. Membandingkan performa dari beberapa algoritma *Machine Learning* dan *Deep Learning* dalam menganalisis sentimen pengguna berdasarkan metrik evaluasi seperti *Accuracy*, *Precision*, *Recall*, dan *F1-Score*, di mana algoritma dengan nilai akurasi tertinggi akan dipilih dan digunakan sebagai model utama.
3. Mengimplementasikan metode *topic modeling* menggunakan algoritma *LDA* guna mengidentifikasi dan mengelompokkan topik-topik utama yang dibahas dalam ulasan pengguna.
4. Mengintegrasikan hasil analisis sentimen dan topik ke dalam visualisasi data yang menggambarkan distribusi sentimen berdasarkan topik dan periode waktu tertentu.
5. Memberikan insight bisnis kepada PT PLN (Persero) terkait persepsi dan tanggapan pelanggan terhadap aplikasi PLN *Mobile* untuk mendukung pengambilan keputusan yang lebih strategis dalam peningkatan layanan.

## 1.5 Batasan Tugas Akhir

Untuk menjaga fokus dan ruang lingkup penelitian agar tetap sesuai dengan kebutuhan perusahaan dan keterbatasan sumber daya yang tersedia, penelitian ini memiliki beberapa batasan sebagai berikut:

Dalam pelaksanaan penelitian ini, terdapat beberapa batasan yang ditetapkan guna menjaga fokus dan ruang lingkup agar tetap terarah serta sesuai dengan tujuan penelitian. Penelitian ini dibatasi hanya pada ulasan pengguna aplikasi PLN *Mobile* yang diambil dari *platform* Google *Play Store* dengan rentang waktu selama satu tahun. Pengambilan data dilakukan menggunakan metode web scraping dan hanya mencakup ulasan yang menggunakan bahasa Indonesia. Ulasan yang menggunakan bahasa lain atau tidak teridentifikasi secara jelas akan dikeluarkan dari proses analisis.

Penelitian ini juga membatasi klasifikasi sentimen ke dalam dua kategori utama, yaitu sentimen positif dan sentimen negatif. Ulasan dengan skor rating 3, 4, dan 5 dikategorikan sebagai sentimen positif, sedangkan ulasan dengan skor 1 dan 2 dikategorikan sebagai sentimen negatif.

Model yang digunakan dalam analisis ini terbatas pada empat algoritma, yaitu *Transformers IndoBERT*, *Naïve Bayes*, *Long Short-Term Memory (LSTM)*, dan *Logistic Regression*. Keempat algoritma tersebut dipilih karena sudah terbukti relevan dalam penelitian-penelitian sebelumnya terkait analisis sentimen. Evaluasi performa dari masing-masing model dilakukan dengan mengacu pada metrik evaluasi utama, yaitu nilai akurasi, *Precision*, *Recall*, dan *F1-Score*. Target minimal yang ditetapkan untuk model terbaik adalah nilai *Recall*  $\geq 86\%$ , *Precision*  $\geq 82\%$ , *F1-Score*  $\geq 84\%$ , dan akurasi  $\geq 86\%$ .

Selain analisis sentimen, penelitian ini juga mengimplementasikan fitur tambahan berupa topik modeling dengan menggunakan metode *LDA* Topik yang dihasilkan dibatasi pada empat kategori utama yang diperoleh dari proses ekstraksi topik pada ulasan pengguna. Visualisasi hasil analisis akan difokuskan pada perbandingan distribusi sentimen berdasarkan topik dan tren bulanan selama tahun 2022.

*Platform* pemrosesan data dilakukan menggunakan dengan bahasa pemrograman *Python*. Penelitian ini tidak mencakup analisis real-time atau data dari *platform* lain di luar *Google Play Store*. Batasan-batasan ini ditetapkan untuk memastikan bahwa penelitian dapat diselesaikan secara terukur dan mendalam sesuai dengan sumber daya dan waktu yang tersedia.