

# BAB I

## PENDAHULUAN

### 1.1. Latar Belakang

PT. Blue Bird Group merupakan perusahaan taksi terbesar yang masih dapat mengembangkan bisnis nya ke berbagai daerah di Indonesia. Kemajuan internet dan perangkatnya yang semakin maju telah mendorong perusahaan taksi besar di Indonesia itu untuk memberikan pengalaman tak terlupakan bagi para penggunanya. Pemesanan taksi kini dipermudah dengan sebuah aplikasi yang menawarkan berbagai kemudahan bagi konsumen. Pengguna smartphone baik itu yang berbasis Android dan iOS dapat mengunduh aplikasi MyBluebird dengan gratis lewat aplikasi *store* masing-masing. Aplikasi MyBluebird merupakan salah satu *Intelligent Transportation System (ITS)* dimana sistem transportasi memanfaatkan kemajuan teknologi. Sebagai pemain lama dalam bidang transportasi darat di Indonesia, PT. Blue Bird Group selalu membuat terobosan baru dalam upaya memberi pelayanan yang maksimal kepada pelanggannya.

Aplikasi MyBluebird pertama kali dirilis untuk pengguna BlackBerry pada Agustus 2011 dan pada bulan Desember 2011 PT. Blue Bird Group merilis untuk pengguna iPhone dan Android. Sebagai bagian dari inovasi bisnis, aplikasi *Taxi Mobile Reservation My Blue Bird* terus dikembangkan. Pada saat ini aplikasi tersebut sudah berganti nama menjadi aplikasi MyBluebird. Pada saat ini aplikasi tersebut terakhir diperbaharui pada tanggal 26 April 2022 menjadi versi 5.19 untuk sistem operasi android dan iOS yang dapat di unduh melalui *Play Store* dan *App Store* serta versi 5.15 pada Huawei dapat diunduh melalui *Huawei AppGalery*. Selain fitur, aplikasi tersebut juga memperbaharui tampilan aplikasi mereka agar mudah digunakan. Fitur lain yang disediakan selain untuk melakukan pemesanan layanan adalah penumpang dapat memilih jenis layanan sesuai dengan kebutuhan. Layanan yang dapat dipesan melalui aplikasi adalah pemesanan

taksi online (Bluebird dan Silverbird), penyewaan rental mobil (Goldenbird), dan layanan pengiriman (Bluebird Kirim). Selain layanan aplikasi ini juga dilengkapi dengan beberapa fitur lainnya seperti chat kepada pengemudi, menandai tempat yang sering dikunjungi, fitur untuk memesan kembali layanan yang pernah digunakan, melihat estimasi pengemudi akan tiba, tidak hanya dalam bentuk argo kini aplikasi tersebut sudah menyediakan harga pasti agar penumpang tidak khawatir akan kekurangan biaya perjalanan, *multi payment* digunakan untuk penumpang menyimpan lebih dari satu jenis pembayaran. Adapun pembayaran dapat dilakukan menggunakan kartu kredit, *e-voucher*, *trip voucher*, *GoPay*, *ShopeePay*, *LinkAja*, dan *i.saku*.

Berdasarkan data di situs Google Play pada tanggal 19 Juli 2022, aplikasi MyBluebird telah diunduh sebanyak lebih dari satu juta dengan rating 4,1 dari total 57,7ribu ulasan. Review pengguna berisi dua bagian, yaitu nilai rating berupa bintang dan komentar secara tekstual. Nilai rating menunjukkan evaluasi keseluruhan pengalaman pengguna menggunakan skala numerik, namun komentar tekstual mampu bercerita lebih mendalam. Review dari pengguna di PlayStore, terdiri dari review positif dan negatif seperti keluhan, kritik, atau saran [1]. Dengan menggunakan data yang tersedia pada kolom ulasan mengenai aplikasi MyBluebird dapat dilakukan analisis pendapat dan opini publik mengenai pengalaman pengguna menggunakan aplikasi MyBluebird dengan cara mengklasifikasi pendapat tersebut menjadi tiga kelas yaitu positif, netral, dan negatif. Penerapannya dengan cara mengambil data melalui kolom ulasan atau komentar pada *google play* mengenai aplikasi MyBluebird kemudian akan diklasifikasikan dengan melihat komentar tersebut merupakan kalimat yang bersifat positif, netral, atau negatif. Metode *Naïve Bayes Classifier* digunakan untuk melakukan analisis sentimen melalui komentar yang diunggah oleh pengguna. Pada sebuah penelitian berjudul Menurut penelitian yang dilakukan oleh Moraes dimana penelitian tersebut membandingkan antara supervised learning dengan *artificial neural network* (ANN) hasilnya adalah

*supervised learning* cenderung kurang terpengaruh oleh istilah *noise* ketika keseimbangan data meningkat, *supervised learning* jauh lebih cepat jika dibandingkan dengan ANN [2]. ANN menggunakan pendekatan BoW dimana metode deep learning ini memiliki beberapa kelemahan, pertama mengabaikan urutan kata, yang berarti bahwa dua dokumen dapat memiliki representasi yang sama persis jika mereka berbagi kata yang sama [1]. Komparasi Akurasi Naïve Bayes dan *Support Vector Machine* (SVM) untuk Rekomendasi Produk in Fashion Dress, menghasilkan kesimpulan bahwa metode Naïve Bayes Classifier memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan *Support Vector Machine* (SVM) [3]. Selain itu berdasarkan penelitian yang dilakukan oleh Mardiana dengan melakukan komparasi antara 5 metode analisis sentimen yaitu *Naïve Bayes Classifier*, *Neural Network*, K-NN, SVM, dan *Decision Tree* menghasilkan bahwa metode *Naïve Bayes Classifier* disarankan untuk digunakan sebagai metode analisis teks, selain mendapatkan nilai yang baik (good classification) pada kelasnya, *Naïve Bayes Classifier* juga terbilang simpel dengan performa yang kuat untuk metode yang mengandalkan asumsi secara keseluruhan [4]. Komentar yang telah pengguna sampaikan pada kolom ulasan kita tidak dapat mengetahui apakah opini tersebut bernilai positif, netral atau bahkan negatif. Oleh karena itu, dengan adanya analisis sentimen kita dapat mengetahui apakah opini yang telah disampaikan oleh pengguna mengenai pengalaman mereka menggunakan aplikasi MyBluebird bernilai positif, netral, atau negatif.

## 1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan diatas, maka rumusan masalah dapat dirangkum sebagai berikut:

1. Bagaimana proses klasifikasi sentimen komentar pengguna mengenai aplikasi MyBluebird yang disampaikan melalui pada kolom ulasan di *Play Store* dengan metode *Naïve Bayes Classifier*?

2. Berapakah tingkat akurasi sentimen sebagai implementasi metode *Naïve Bayes Classifier*?

### 1.3. Batasan Penelitian

Adapun batasan masalah pada penelitian ini yaitu:

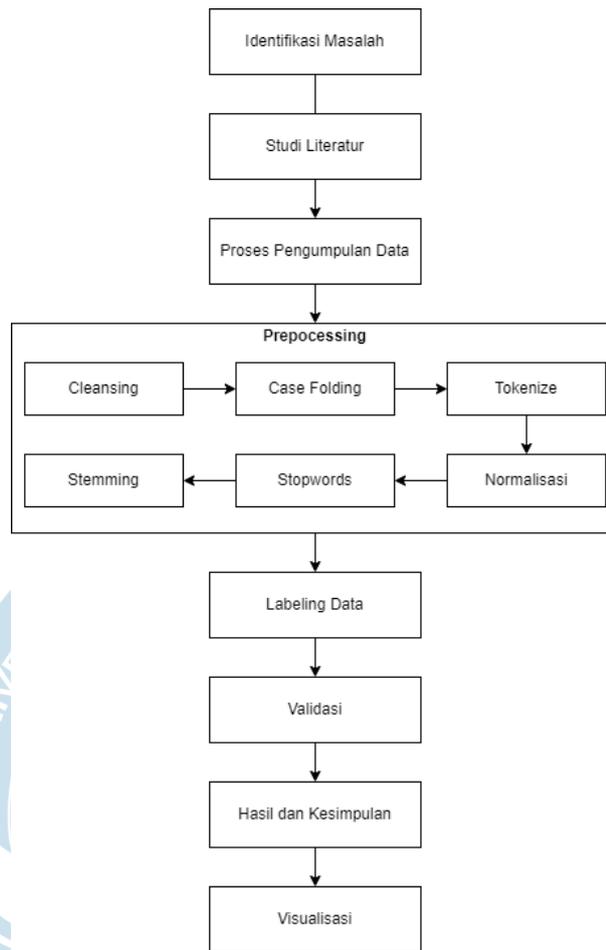
1. Data set yang digunakan hanya diambil pada kolom ulasan aplikasi di *Play Store*.
2. Data set diambil dalam rentan waktu Januari 2021 – 15 Juli 2022.
3. Data yang digunakan adalah komentar yang diunggah oleh pengguna di *Play Store*.
4. Hasil akhir dari penelitian ini adalah analisis sentimen dari metode *Naïve Bayes Classifier*.
5. Menggunakan algoritma pemrograman *python* sebagai *collecting data*.

### 1.4. Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah penulis ingin menghasilkan sebuah nilai sentimen dari hasil analisa komentar pengguna mengenai pengalaman mereka menggunakan aplikasi MyBluebird yang diunggah pada kolom ulasan di *Play Store*. Dengan menggunakan metode *Naïve Bayes Classifier* penulis ingin melakukan klasifikasi serta mengetahui tingkat akurasi dari penggunaan metode tersebut.

### 1.5. Metode Penelitian

Metode penelitian adalah suatu proses berpikir guna memastikan permasalahan, mengumpulkan data baik melalui studi literatur ataupun melalui penelitian lapangan, melaksanakan pengolahan data hingga memberikan kesimpulan dari permasalahan yang diteliti. Dengan metodologi sehingga siklus pemecahan permasalahan pada penelitian dapat dilaksanakan secara baik serta terstruktur. Metode penelitian yang hendak digunakan pada penelitian ini sebagaimana ditunjukkan pada **Gambar 1.1**



**Gambar 1.1** Alur Penelitian

### 1.5.1. Identifikasi Masalah

Penelitian ini menggunakan metode *Naïve Bayes Classifier* dengan studi kasus aplikasi MyBluebird sebagai acuan penelitian. Data diambil berdasarkan komentar pengguna yang disampaikan melalui kolom ulasan aplikasi yang tersedia di Google Play Store. Data akan diambil dan diklasifikasi menjadi tiga kelas yaitu sentimen positif, sentimen netral, dan sentimen negatif. Setelah itu data akan divalidasi menggunakan metode *k-fold cross validation* untuk menentukan nilai akurasi.

### 1.5.2. Studi Literatur

Studi ini dilakukan dengan cara mempelajari, meneliti dan menelaah berbagai literatur yang bersumber dari buku-buku, teks, jurnal ilmiah, situs-situs di internet, dan bacaan yang terkait dengan topik analisis sentimen berdasarkan aspek, metode preprocessing, metode TF-IDF, dan metode klasifikasi *Naïve Bayes Classifier*.

### 1.5.3. Proses Pengumpulan Data

Tahapan pengumpulan data merupakan proses yang penting dari suatu penelitian. Data penelitian yang digunakan harus akurat dan jelas sumbernya. Pengumpulan data yang digunakan dalam penelitian diambil dari kolom ulasan aplikasi MyBluebird yang tersedia di *Google Play Store* melalui library *google-play-scraper* dengan menggunakan Python yang kemudian dipisahkan menjadi sentimen positif, sentimen netral, dan sentimen negatif. Sentimen positif yaitu sentimen dimana pengguna memberikan komentarnya dengan memberikan opini mereka secara positif atau memberikan keterangan yang menunjukkan kepuasan saat menggunakan aplikasi MyBluebird. Sentimen netral yaitu sentimen dimana pengguna tidak menunjukkan kepuasan serta juga tidak menunjukkan ketidakpuasan saat menggunakan aplikasi MyBluebird. Sedangkan sentimen negatif merupakan sentimen sebaliknya dari sentimen positif yaitu pengguna merasakan ketidaknyamanan saat menggunakan aplikasi MyBluebird.

- Analisis Data

Pada bagian ini tiap data digabung sebagai satu dataset. Dari dataset yang telah ada setelah itu diidentifikasi setiap fitur yang terdapat pada data, dimana fitur merupakan subset dari atribut yang digunakan pada model

#### 1.5.4. Preprocessing Data

*Preprocessing* atau yang bisa disebut dengan tahap persiapan awal data, dimana komentar public yang didapatkan pada kolom ulasan aplikasi pada Play Store dilakukan beberapa standarisasi dikarenakan komentar seseorang dituliskan dengan kata atau kalimat yang tidak terstruktur. Beberapa tahap yang terjadi dalam *preprocessing* data adalah:

- *Cleansing*, merupakan tahap penghapusan noise yang terdapat pada data seperti tanda baca, simbol, emoticon maupun angka.
- *Case Folding*, merupakan tahap menjadikan semua data memiliki format *lowercase* atau huruf kecil semua.
- *Tokenizing*, merupakan proses pemotongan teks menjadi per-kata.
- *Normalisasi*, merupakan proses menerjemahkan kata-kata gaul atau tidak baku ke kata terdekat dengan menggunakan kamus yang dibuat untuk melihat pola kemunculan kata-kata tidak baku tersebut
- *Stopword*, digunakan untuk membuang kata-kata yang sering muncul dan bersifat umum, kurang menunjukkan relevansinya dengan teks.
- *Stemming*, merupakan proses mencari akar kata dan menghilangkan imbuhan pada kata. Tujuannya mengurangi variasi kata yang memiliki kata dasar sama.

#### 1.5.5. Labeling Data

Tahapan labeling data dilakukan setelah data yang telah dikumpulkan sudah melalui tahap *preprocessing*. Dimana data yang sudah bersih atau siap digunakan untuk penelitian dilakukan labeling berdasarkan pada pengetahuan domain individu dan pemahaman bahasa [24]. Labeling data terdiri dari label positif, netral, dan negatif. Dalam

menentukan sebuah kalimat apakah termasuk sentimen positif, negatif, atau netral penulis menggunakan Kamus Besar Bahasa Indonesia (KBBI) sebagai acuan untuk melakukan labeling untuk menentukan sebuah kalimat mengandung kata sentimen positif atau negatif. Proses labeling data dilakukan secara manual di *Microsoft Excel*.

#### **1.5.6. Validasi**

Setelah melalui tahap labeling data, selanjutnya data akan dibagi menjadi 2 yaitu data training dan data testing. Dimana data tersebut akan dilakukan penelitian menggunakan metode *Naïve Bayes Classifier*. Setelah itu, data akan divalidasi menggunakan metode *k-fold cross validation* dan dilakukan sebanyak 10 iterasi.

#### **1.5.7. Hasil dan Penarikan Kesimpulan**

Pada penarikan kesimpulan terakhir sesudah data diuji, sehingga kita dapat mengambil kesimpulan dari hasil uji mengenai tiap data dalam bentuk persentasi yang menggambarkan data tersebut. Hasil kesimpulan akan dapat diketahui data cenderung positif, netral, atau negatif.

#### **1.5.8. Visualisasi data dengan *Word Cloud* dan *Diagram Pie***

Komentar yang telah disampaikan pengguna melalui kolom ulaan aplikasi di Play Store dapat divisualisasikan dalam bentuk *Word Cloud* dan *Diagram Pie*. *Word Cloud* adalah representasi grafis dari sebuah dokumen yang dilakukan dengan memplot kata-kata yang sering muncul dalam dokumen dalam ruang dua dimensi. Frekuensi kata yang sering muncul ditunjukkan dengan ukuran huruf dari kata tersebut. Semakin besar ukuran kata, semakin besar frekuensi kemunculannya dalam dokumen. *Diagram pie* menunjukkan persentase dari hasil analisis sentimen. Memudahkan pembaca guna mengetahui seberapa banyak komentar positif, netral, dan negatif.

### **1.5.9. Sistematika Penulisan**

#### **BAB I PENDAHULUAN**

Pada bab ini berisikan konsep dasar dari pembuatan dan penulisan laporan tugas akhir dimana pada bab ini berisikan latar belakang, rumusan masalah, batasan penelitian, tujuan penelitian, metode penelitian, dan sistematika penulisan.

#### **BAB II TINJAUAN PUSTAKA**

Pada bab ini berisikan beberapa penelitian serupa yang telah terlaksana, dimana penelitian tersebut menjadi acuan untuk penelitian yang sedang dilakukan.

#### **BAB III LANDASAN TEORI**

Pada bab ini berisikan penjelasan berupa teori-teori pendukung dari penelitian yang akan dilakukan oleh penulis.

#### **BAB IV ANALISIS PERANCANGAN EKSPERIMEN**

Pada bab ini berisikan identifikasi masalah, studi literatur, proses pengumpulan data, preprocessing data, proses labeling data, proses persiapan dataset serta hasil dan kesimpulan.

#### **BAB V HASIL ANALISA DAN PEMBAHASAN**

Pada bab ini berisikan deskripsi eksperimen, hasil eksperimen, dan pembahasan eksperimen.

#### **BAB VI PENUTUP**

Pada bab ini berisikan kesimpulan dari penelitian yang telah dilakukan, serta saran yang didapatkan penulis yang dapat digunakan untuk penelitian selanjutnya.