

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Ulasan atau *Review* adalah sebuah pendapat yang diuraikan oleh pengguna atau konsumen yang telah menggunakan produk tersebut. Melalui review inilah salah satu cara calon pengguna untuk menentukan apakah produk tersebut bagus atau tidak. Terutama *review* pada aplikasi atau *game* di Platform *google play store* yang memiliki ulasan dengan rating satu sampai lima dan ditambah ulasan teksnya [1]. Analisis sentimen akan menjadi berguna saat pihak dari pengembang *game* ingin mengetahui opini pengguna tentang pengalaman bermain dari gamenya. Dengan membaca ulasan dari platform tersebut merupakan cara yang paling bagus untuk memutuskan arah pengembangan dan peningkatan selanjutnya dari *game* mereka [2]. Analisis sentimen dapat digunakan untuk mengetahui polaritas dari teks ulasan dari *game* tersebut. Polaritas memiliki makna apakah teks ulasan tersebut positif atau negatif [3]. Maka, Dengan menggunakan analisis sentimen untuk menganalisis ulasan tersebut, kita dapat menghasilkan dokumen khusus dan terperinci yang terdapat ulasan pengguna. Dengan memiliki data inilah, pengembangan *game* dapat memperoleh informasi tersebut sebagai suatu keuntungan [4].

Namun terdapat masalah apabila ulasan-ulasan tersebut memiliki lebih dari satu aspek pembahasan dengan polaritas yang berbeda sehingga membuat sebuah kalimat ulasan tidak hanya memiliki satu jenis polaritas saja [5]. Terutama pada ulasan-ulasan dari sebuah *Game/gim* yang bernama *Genshin Impact*, Tidak hanya itu, banyaknya jumlah sentimen yang terdapat pada ulasan-ulasan tersebut tidak selalu sama dengan ulasan lainnya sehingga akan menyebabkan ketidaksamaan jumlah data yang terdistribusi pada setiap kelasnya. Masalah ini sering ditemui pada kondisi nyata. Kondisi ini dinamakan ketidakseimbangan data atau *imbalanced data*. Dimana, salah satu

kelasnya memiliki jumlah yang sangat besar dibandingkan dengan kelas lainnya [6] maka dari itu, akan diperlukannya penanganan untuk *imbalanced data* ini agar saat penentuan klasifikasi tidak mengabaikan kelas minoritas akibat dominasi kelas mayoritas.

Penelitian sebelumnya yang berjudul “Deep Learning and Data Sampling with Imbalanced Big Data” oleh Justin M. Johnson dan Taghi M. Khoshgoftaar [7] menggunakan *Deep Learning* dengan metode sampling *Random Under Sampling*, *Random Over Sampling*, dan *Random Over-Under Sampling* pada Big Data yang tidak seimbang, hasilnya ROS sedikit unggul dari RUS serta ROUS mampu memaksimalkan performa dan efisiensi dengan mempercepat waktu proses training hingga 4 kali lipat. Kemudian, Pada penelitian Edyt Daryfayi Putra Daulay dan Ibnu Asror [1] yang berjudul “Sentimen Analisis Pada Ulasan Google Play Store Menggunakan Metode Naïve Bayes” telah menghasilkan akurasi cukup bagus yakni 78,9% dengan metode MNB sebagai model klasifikasi sentimen untuk kasus klasifikasi ulasan berdasarkan teks. Ini juga didukung oleh M. Ataur Rahman dan Y. Ara Akter [8] yang melakukan penelitian berjudul “Topic Classification from Text Using Decision Tree, K-NN and Multinomial Naïve Bayes” dimana mereka memilih Multinomial Naïve Bayes sebagai model klasifikasi terbaik yang mereka gunakan.

Berdasarkan latar belakang dan uraian diatas, maka pada penelitian ini akan dilakukan penanganan data imbalance dengan RUS (*Random Under Sampling*) dan ROS (*Random Over Sampling*) pada Analisis Sentimen Berbasis Aspek (ABSA) menggunakan metode Naïve Bayes Multinomial sebagai klasifikasi sentiment positif dan negatif dari masing masing aspek dan melihat bagaimana pengaruhnya pada performa klasifikasi Algoritma *Naïve Bayes Multinomial* dalam melakukan Klasifikasi terhadap Ulasan gim *Genshin Impact* yang dimana MNB merupakan *Supervised Learning*. Maka telah diputuskan bahwa Judul penelitian ini adalah “Pengaruh Penanganan Data Imbalance Pada Analisis Sentimen Berbasis Aspek Terhadap Review Game Genshin Impact Dengan Naïve Bayes Multinomial”.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas maka rumusan masalah dari penelitian ini adalah sebagai berikut.

1. Bagaimana Performa algoritma Naïve Bayes Multinomial dalam menganalisis sentimen teks ulasan berbahasa Indonesia pada game Genshin Impact berdasarkan aspek?
2. Bagaimana pengaruh dari teknik balancing data RUS (*Random Under Sampling*) dan ROS (*Random Over Sampling*) pada performa algoritma Naïve Bayes Multinomial dalam mengklasifikasi sentimen berbasis aspek pada ulasan *game Genshin Impact* berbahasa Indonesia?

1.3 Tujuan Penelitian

Berdasarkan Rumusan masalah diatas maka Tujuan penelitian ini adalah:

1. Mengetahui Performa algoritma Naïve Bayes Multinomial dalam menganalisis sentimen teks ulasan berbahasa Indonesia pada game Genshin Impact berdasarkan aspek.
2. Mengetahui pengaruh dari teknik balancing data RUS (*Random Under Sampling*) dan ROS (*Random Over Sampling*) pada performa algoritma Naïve Bayes Multinomial dalam mengklasifikasi sentimen berbasis aspek pada ulasan Genshin Impact.

1.4 Batasan Masalah

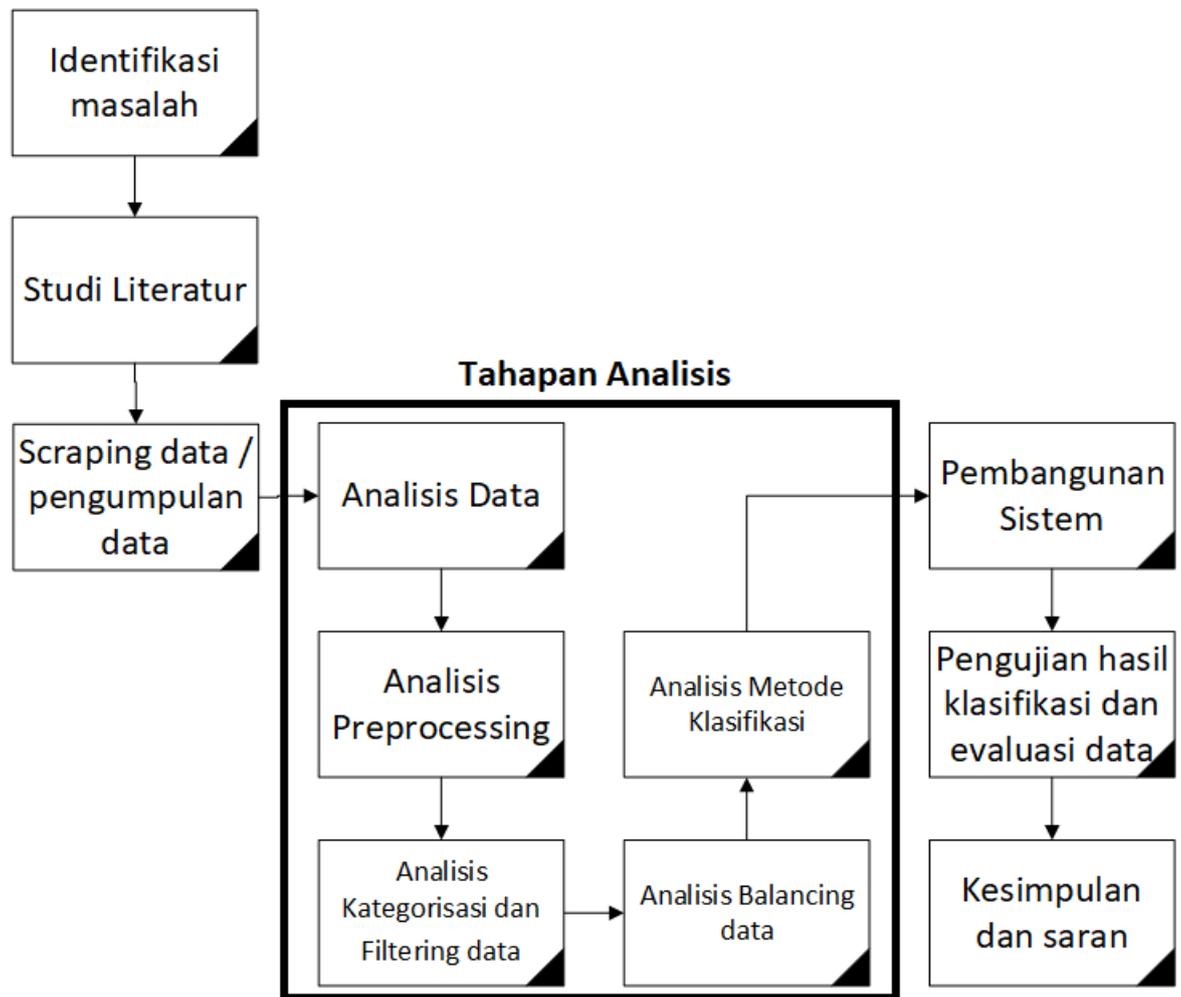
Berikut batasan-batasan masalah pada penelitian ini.

1. Data yang dianalisa adalah data hasil scraping dari review *Genshin Impact* pada *Google Play Store* Berbahasa Indonesia
2. Data Scraping dari tanggal 21/07/2021
3. Penentuan aspek berdasarkan hasil pengamatan dataset, aspek-aspek tersebut adalah: aspek Gameplay dan Storyline, Aksesibilitas, Grafis dan Performa, dan Fitur.
4. Jumlah kelas terbagi 2 yaitu positif dan negatif.
5. Jumlah data valid yang akan digunakan sebanyak 1000 review.

6. Isi review yang tidak memiliki aspek akan dibuang, contoh: “nvm...”, “no comen”, “bagussssssss”, “Bismillah gacha wany wany”.

1.5 Metode Penelitian

Metodologi penelitian adalah suatu proses yang dipakai untuk memecahkan permasalahan yang logis. Proses ini memerlukan data-data untuk mendukung berjalannya suatu penelitian. Metode penelitian yang digunakan pada penelitian ini menggunakan metode Penelitian Kuantitatif, dimana metode ini menguji teori tertentu dengan meneliti antar variabel yang diukur dengan instrument penelitian sehingga data yang terdiri dari angka-angka dapat dianalisis berdasar prosedur statistik [9]. Berikut ini adalah tahapan penelitian yang akan dilakukan dengan alur penelitian seperti pada Gambar 1 dibawah ini.



Gambar 1 Alur Penelitian

1.5.1 Identifikasi Masalah

Tahap awal yang dilakukan pada penelitian ini adalah Identifikasi Masalah. Dimana, Ulasan-ulasan game *Genshin impact* di *Google playstore* memiliki lebih dari satu aspek bahasan dengan polaritas yang berbeda-beda di satu ulasan yang sama dan juga ulasan-ulasan tersebut tidak seimbang.

1.5.2 Studi Literatur

Studi literatur merupakan serangkaian metode pengumpulan referensi jurnal, paper, dan buku yang memiliki keterkaitan dengan analisis sentimen berbasis aspek.

1.5.3 Pengumpulan dataset

Data sampel atau *dataset* yang digunakan pada penelitian ini berupa Review terhadap game *Genshin Impact* di *Google Playstore* dengan cara melakukan Teknik scraping menggunakan pemrograman *python*.

1.5.4 Tahapan Analisis

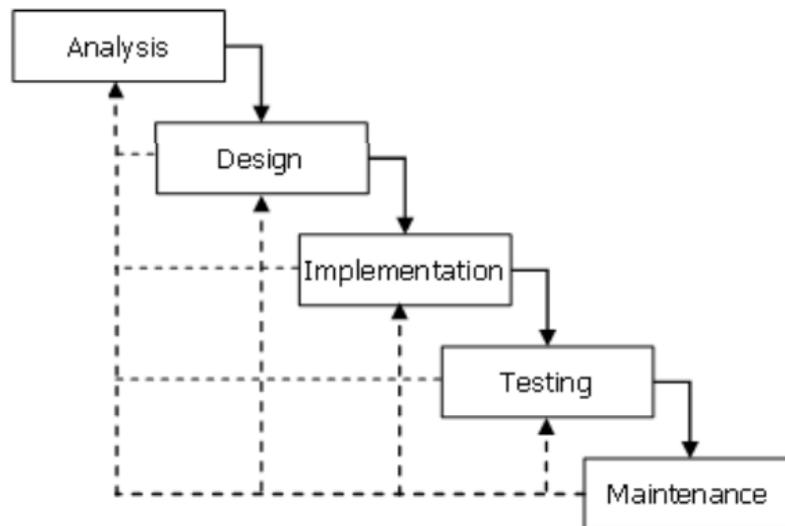
Setelah mendapatkan pemahaman dari Studi literatur seperti membaca jurnal, paper, dan buku, maka tahapan analisis pada penelitian ini adalah:

1. Analisis Data, yaitu mengamati isi dataset yang didapatkan yang kemudian menentukan aspek-aspek yang sesuai dengan hasil pengamatan. Aspek-aspek tersebut yaitu Aspek Gameplay dan Story, Aspek Aksesibilitas, Aspek Grafis dan Performa, dan Aspek Fitur. Data-data yang dianalisis dan diamati isinya adalah kalimat dari ulasan *Genshin Impact* pada *Google playstore*.
2. Analisis *Pre-processing*, Pada tahap ini, kalimat ulasan dari *dataset* yang terkumpul akan diproses menjadi sebuah data terstruktur yang mudah untuk diolah. Langkah-langkah *pre-processing* yang dilakukan terdiri dari *case folding*, *cleansing*, *normalisasi*, *stopword*, *tokenizing*, dan *stemming*.
3. Analisis Kategori aspek dan Filtering data, Analisis Kategori Aspek adalah tahapan dimana data ulasan akan diberikan label aspek sesuai dengan aspek yang terkandung didalamnya. Dan Filtering data adalah tahapan setelah tahap kategori aspek, dimana sistem akan membuang ulasan yang tidak layak.
4. Analisis Data Balancing, dikarenakan data sampel memiliki label positif dan negatif yang tidak seimbang, maka diperlukannya menyeimbangkan data. Teknik menyeimbangkan data yang digunakan pada penelitian ini adalah *Random Under Sampling* (RUS) dan *Random Over Sampling* (ROS).
5. Analisis Metode Klasifikasi, pada tahap ini akan dilakukan analisis klasifikasi menggunakan metode *Naïve Bayes Multinomial* meliputi perhitungan probabilitas kelas positif dan negatif, probabilitas atribut

yang terdapat pada ulasan berdasarkan kelas positif dan negatif, perhitungan akhir probabilitas kelas dan probabilitas atribut berdasarkan kelas, dan perbandingan hasil untuk menentukan hasil klasifikasi.

1.5.5 Pembangunan Sistem

Setelah tahapan analisis, selanjutnya adalah untuk Pembangunan sistem atau perangkat lunak. Model pembangunan sistem atau perangkat lunak menggunakan model waterfall [10]. Model waterfall digunakan karena waterfall mendefinisikan tahapan yang secara berturut turut harus diselesaikan satu per satu dan akan lanjut ke tahap selanjutnya jika tahapan sebelumnya telah selesai. Dan juga waterfall bersifat rekrusif dimana seluruh tahapan dapat diulang tanpa batas hingga tahapan itu telah disempurnakan. Berikut dibawah ini merupakan gambaran dari model waterfall.



Gambar 2 Gambaran Model Waterfall

Berikut dibawah ini merupakan penjelasan tahapan waterfall

1. Analysis

Sering dikenal sebagai Software Requirement Specification (SRS) merupakan penjelasan lengkap tentang sifat dari sistem atau perangkat

lunak yang akan dibangun dengan mendefinisikan kebutuhan non-fungsional dan fungsional.

2. Design

Tahapan ini merupakan tahap pembangunan desain sistem atau perangkat lunak sesuai dengan kebutuhan yang ada pada tahap *analysis*.

3. Implementation

Tahapan ini merupakan tahapan implementasi hasil rancangan dari *design*. Implementasi dilakukan menggunakan pemrograman PHP dan javascript berbasis website dengan penyimpanan data menggunakan MySQL. Dan untuk Web Servernya menggunakan sebuah software berbasis web server yaitu Xampp

4. Testing

Tahapan ini merupakan tahapan setelah tahap implementasi untuk menguji sejauh mana sistem dapat bekerja dengan baik.

5. Maintenance

Tahapan ini merupakan tahapan terakhir dimana pada tahap ini akan dilakukan proses memodifikasi solusi perangkat lunak, memperbaiki keluaran, memperbaiki kesalahan, dan meningkatkan kualitas dan kinerja dari sistem atau perangkat lunak agar dapat beradaptasi pada lingkungan bila diperlukan.

1.5.6 Pengujian

Dalam tahapan ini, akan dilakukan pengujian Akurasi dan F1-Score dari metode *Multinomial Naïve Bayes* terhadap analisis sentimen berbasis aspek pada Review game *Genshin Impact*. Pengujian juga dilakukan pada masing masing dataset yaitu Data non-balance, data Undersampled, dan data Oversampled.

1.5.7 Kesimpulan dan Saran

Pada tahapan ini akan dilakukan penarikan kesimpulan dan saran berdasarkan hasil dari tahap pengujian dalam menerapkan Penanganan data

imbalance dengan Metode *Naïve Bayes Multinomial* pada kasus analisis sentiment berbasis aspek pada review game *Genshin Impact*.

1.6 Sistematika penulisan

Sistematika penulisan disusun untuk memberikan gambaran umum tentang penulisan tugas akhir yang akan dilakukan. Sistematika penulisan tugas akhir ini adalah sebagai berikut:

BAB 1 PENDAHULUAN

Pada Bab 1 akan dijelaskan tentang latar belakang masalah yang diambil, identifikasi masalah, maksud dan tujuan, batasan masalah, metodologi penelitian dan sistematika penulisan.

BAB 2 TINJAUAN PUSTAKA

Pada Bab 2 akan dijelaskan tentang tinjauan umum pembahasan berbagai konsep dasar mengenai Text Mining, Scraping, Analisis Sentimen, Analisis Sentimen Berbasis Aspek, Text Pre-processing, case folding, cleansing, normalization, stop word, tokenizing, stemming, Naïve Bayes, Naïve Bayes Multinomial, Laplace Smoothing, Data Balancing, Random Undersampling, Random Oversampling, Confusion Matrix dan teori-teori pendukung lainnya yang berkaitan dengan topik pembangunan perangkat lunak.

BAB 3 ANALISIS DAN PERANCANGAN

Pada Bab 3 ini berisi tentang analisis dan perancangan aplikasi yang dibangun yang meliputi analisis masalah, analisis data, analisis sistem, analisis preprocessing, analisis balancing data, analisis metode Naïve Bayes Multinomial, UML dan Perancangan antarmuka.

BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM

Pada Bab 4 ini berisi tentang gambaran dari hasil implemetasi dan pengujian pada sistem yang mulai pada implementasi sistem, pengujian sistem dengan black box dan confusion matrix dari proses pengklasifikasian oleh metode Naïve Bayes Multinomial dengan menampilkan hasil keluaran seperti nilai akurasi dan F1-Score dari analisis sentiment berdasarkan aspek terhadap review game *Genshin Impact*.

BAB 5 KESIMPULAN DAN SARAN

Pada Bab 5 akan dijelaskan tentang kesimpulan yang sudah diperoleh dari hasil penulisan tugas akhir dan saran agar dapat dilakukan pengembangan untuk penelitian selanjutnya