

# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Dalam penelitian sentimen yang berfokus pada aspek-aspek tertentu, permasalahan utama yang dihadapi adalah ketidakseimbangan antar kelas sentimen. Ketidakseimbangan terjadi ketika suatu kelas memiliki jumlah data yang lebih banyak daripada kelas lainnya [1]. Pada kasus pembatalan piala dunia U-20 di Indonesia keputusan FIFA untuk mencabut status menjadi perbincangan menarik di media sosial, terutama di Twitter, di mana berbagai tanggapan positif dan negatif bermunculan [13]. Dampak dari ketidakseimbangan kelas sentimen ini dapat menghasilkan performa model yang kurang baik. Namun, sering kali dalam sebuah dataset terjadi ketidakseimbangan data antara jumlah sampel yang mewakili berbagai kelas sentimen. Pada penelitian ini menggunakan random over sampling (ROS) digunakan untuk menangani ketidakseimbangan kelas sentimen berbasis aspek dengan membuat contoh kelas minoritas sintetis menjadi sama dengan kelas jumlah mayoritas sintetis.

Pada penelitian sebelumnya yang dilakukan oleh Johnson dan Khoshgoftaar [25] mengatasi ketidakseimbangan data dari Center for Medicare and Medicaid Services (CMS) menggunakan metode random undersampling (RUS), random oversampling (ROS), dan kombinasi random over-under sampling. Hasil penelitian menunjukkan bahwa ROS sedikit lebih unggul dibandingkan RUS dan kombinasi ROUS. Metode ROS mampu memaksimalkan kinerja dan efisiensi dengan mempercepat proses pelatihan hingga empat kali lipat.

Pada penelitian selanjutnya dilakukan oleh Afifah Nurlaila dkk [3] yang menggunakan ROS dalam menangani kelas yang tidak seimbang, dalam penelitian tersebut algoritma resampling ROS dapat meningkatkan nilai recall, precision, dan F1-measure. Setelah menggunakan algoritma resampling ROS, rata-rata nilai precision, recall, dan F1-Measure tertinggi terdapat pada kelas no-emotions yaitu precision bernilai 96,64%, recall bernilai 76,36%, dan F1-Measure bernilai 85,93%.

Kemudian penelitian yang dilakukan oleh Resianta Perangin-angin dkk [1] dalam kasus ketidakseimbangan kelas dalam penggunaan algoritma ROS dan random under sampling (RUS) dapat mengatasi tingkat ketidakseimbangan data menjadi lebih baik, untuk kedua algoritma tersebut dari ujicoba yang

dilakukan didapat bahwa algoritma ROS memiliki kemampuan yang lebih baik untuk menyeimbangkan data dibandingkan dengan algoritma RUS dalam 5 dataset yang di ujicobakan.

Pada penelitian yang dilakukan oleh Nur Mutmainnah Djafar dan Achmad Fauzan [26] yang menggunakan ROS, Adaptive Synthetic (ADASYN), Synthetic Minority Oversampling Technique (SMOTE) dan algoritma K-Nearest Neighbour (KNN) untuk kasus ketidakseimbangan kelas ditemukan bahwa dari ketiga teknik tersebut, ROS adalah yang paling efisien dengan  $k=5$ , menghasilkan kinerja terbaik dalam hal sensitivitas, spesifisitas, G-mean, dan akurasi masing-masing mencapai 0,643, 0,805, 0,719, dan 78,873%.

Berdasarkan masalah-masalah yang telah dipaparkan diatas oleh beberapa peneliti sebelumnya maka penerapan pada penelitian ini akan dilakukan penerapan optimasi metode ROS untuk kasus imbalance class dalam analisis sentimen berbasis aspek.

## **1.2 Rumusan Masalah**

Berdasarkan latar belakang yang telah dipaparkan maka dirumuskan sebuah permasalahan yaitu bagaimana mengukur performansi dari ROS dalam melakukan analisis sentimen berbasis aspek pada kasus pembatalan piala dunia U-20 di Indonesia.

## **1.3 Maksud**

Menerapkan metode ROS secara efektif dalam konteks analisis sentimen berbasis aspek pada kasus pembatalan piala dunia U-20 di Indonesia.

## **1.4 Tujuan**

Mengukur performansi dari metode ROS dalam melakukan analisis sentimen berbasis aspek pada kasus pembatalan piala dunia U-20 di Indonesia.

## **1.5 Batasan Masalah**

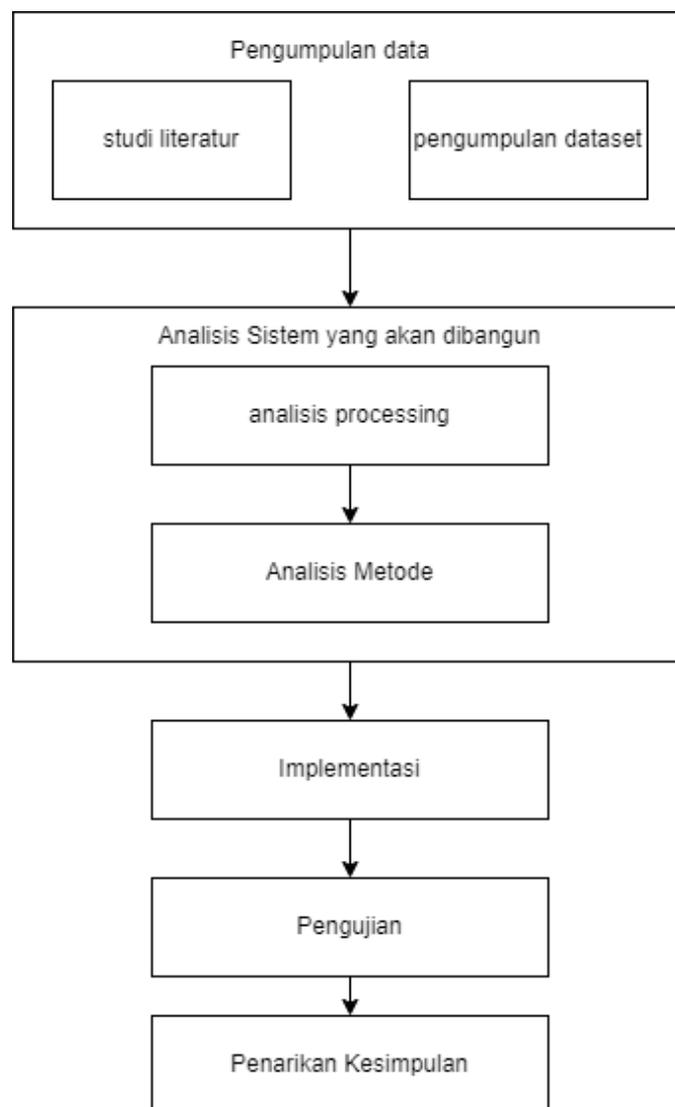
Adapun batasan masalah yang diteliti, supaya lebih terarah dan dapat dipahami maka penelitian ini diberikan batasan masalah, yaitu :

1. Data yang diambil merupakan bahasa indonesia
2. Menggunakan ekstraksi fitur TF-IDF
3. Metode klasifikasi yang digunakan yaitu metode K-NN
4. Metode penanganan imbalance data menggunakan ROS
5. Aspek yang akan dipertimbangkan meliputi politik, sosial, ekonomi, dan olahraga
6. Polaritas yang telah diproses akan diklasifikasikan menjadi kelas positif, negatif, dan netral
7. Penelitian ini akan menerapkan analisis sentimen dengan menggunakan metode multilabel.
8. Data yang diambil melalui proses scrapping pada dataset aplikasi X (twitter)
9. Hastag yang digunakan yaitu #PialaDuniaU20 dan #PialaDuniaU20Indonesia
10. Kata pencarian yang digunakan yaitu fifa batal u-20
11. Pelabelan sentimen dilakukan secara manual
12. Format file untuk data latih dan data uji adalah .csv

## **1.6 Metodologi Penelitian**

Metode penelitian yang digunakan pada penelitian ini adalah metode penelitian kuantitatif. Tahapan awal dalam penelitian ini dirumuskan masalah yang dibahas dalam penelitian ini. Untuk menyelesaikan masalah yang sudah dirumuskan peneliti melakukan studi Pustaka untuk mencari solusi yang terbaik untuk menyelesaikan masalah. Untuk menguji solusi hasil studi literatur peneliti mengumpulkan data yang akan diproses. Kemudian mencoba menganalisis data dan metode yang digunakan. Hasil dari analisis,

diimplementasikan dengan membangun sebuah simulator. Untuk membuktikan teori dan implementasi berjalan sesuai dilakukan pengujian dan evaluasi. Penarikan kesimpulan adalah tahap akhir dalam penelitian ini. Adapun tahapan dari metode penelitian yang digunakan dapat dilihat pada gambar 1.6..



**Gambar 1. 1 Metodologi Penelitian**

### 1.6.1 Metode Pengumpulan Data

Berikut ini merupakan metode pengumpulan data yang akan dilakukan pada penelitian ini adalah sebagai berikut:

#### 1. Studi Literatur

Penulis melakukan studi literatur dengan cara mempelajari sumber-sumber informasi dari jurnal ilmiah, paper, buku dan konten-konten yang tersedia di internet dalam bentuk teks artikel.

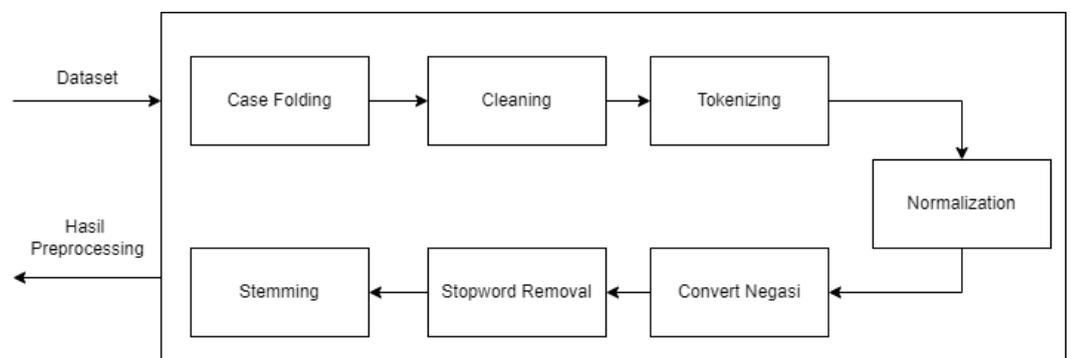
#### 2. Pengumpulan Dataset

Dataset yang digunakan adalah data yang diambil melalui proses scrapping pada aplikasi X(twitter).

### 1.6.2 Metode Tahapan Analisis

Tahapan Analisis pada penelitian ini sebagai berikut:

1. Analisis Preprocessing, merupakan tahap untuk memproses data agar data menjadi lebih terstruktur dan lebih siap untuk diolah. Langkah yang dilakukan pada tahap preprocessing terdiri dari Case Folding, Cleaning, Normalization, Stopword, Tokenization, Stemming, Convert Negasi



*Gambar 1. 2 Analisis Preprocessing*

2. Analisis Metode, pada tahap ini akan dilakukan klasifikasi dengan menggunakan metode random over sampling (ROS).

### **1.6.3 Implementasi**

Pada tahapan implementasi akan dilakukan pembuatan sistem berupa simulator, adapun hal yang bisa dilakukan oleh simulator yaitu memasukkan dan membaca data ulasan yang telah dilabeli secara manual, kemudian melakukan implementasi preprocessing yaitu: case folding, cleaning, normalization, stopword, tokenizing, stemming dan convert negation. Setelah data masukan dipraproses, tahap selanjutnya yaitu pembobotan kata dengan menggunakan Term Frequency-Inverse Document Frequency (TF-IDF).

Lalu data dari seleksi fitur akan dilakukan pembagian data, yaitu data latih dan data uji. Lalu implementasi algoritma dengan menggunakan ROS yang akan menghasilkan kelas.

### **1.6.4 Pengujian**

Dalam tahap ini akan dilakukan pengujian dan evaluasi performansi dari penerapan optimasi metode random over sampling (ROS) untuk kasus imbalance class dalam analisis sentimen berbasis aspek.

### **1.6.5 Penarikan Kesimpulan**

Pada tahap ini akan dilakukan penarikan kesimpulan berdasarkan hasil dari tahap pengujian dalam menerapkan metode random over sampling (ROS) dalam analisi sentimen berbasis aspek.

## **1.7 Sistematika Penulisan**

### **BAB 1 PENDAHULUAN**

Bab ini menguraikan berbagai hal yang berkaitan dengan permasalahan dimana pada penelitian ini, menjelaskan latar belakang penelitian,

rumusan masalah dari penelitian, tujuan dan manfaat penelitian, batasan – batasan penelitian serta sistematika penulisan dari tulisan penelitian ini.

## **BAB 2 LANDASAN TEORI**

Bab ini menguraikan tinjauan umum dilakukannya penelitian. Mulai dari landasan teori yang berkaitan dengan topik penelitian berupa bahan kajian, konsep dasar, dan teori dari para ahli, serta tinjauan terhadap penelitian sebelumnya.

## **BAB 3 ANALISIS DAN PERANCANGAN**

Bab ini menguraikan mengenai analisis serta perancangan sistem yang akan dibangun, mulai dari awal mula hingga hasil akhir sistem untuk menghasilkan sebuah keputusan untuk mengklasifikasikan aspek dan sentimen dari data yang ada. Mencakup mulai dari data mentah, tahapan – tahapan *preprocessing* (*cleaning, case folding, tokenizing, normalization, stemming, dan stopword removal*), tahapan pembobotan kata menggunakan *tf-idf*, dan pembahasan yang mendalam mengenai algoritma yang digunakan beserta perhitungannya sehingga memunculkan hasil akhir.

## **BAB 4 IMPLEMENTASI DAN PENGUJIAN**

Bab ini akan menguraikan tentang implementasi dari sistem yang telah dibangun. Selain itu, bab ini akan melakukan perhitungan akurasi dari hasil pengujian yang telah dilakukan.

## **BAB 5 KESIMPULAN DAN SARAN**

Bab ini merupakan penutup dari sistematika penulisan penelitian, yang berisi ringkasan kesimpulan yang dapat diambil dari seluruh penelitian, serta saran-saran yang diperoleh setelah penelitian dilakukan. Diharapkan bahwa saran-saran ini dapat memberikan masukan yang bermanfaat dan membangun.