

1. Pendahuluan

Pada bab ini akan menjelaskan tentang latar belakang dari penelitian ini, serta akan menjelaskan rumusan masalah, tujuan dan juga sistematika penulisan dalam penelitian ini.

1.1. Latar Belakang

Saat sekarang ini sangat banyak karya ilmiah yang dipublikasikan di seminar nasional maupun internasional. Karya ilmiah merupakan sebuah hasil penelitian maupun gagasan baru. Penelitian tersebut harus didukung oleh karya ilmiah yang sudah dipublikasikan sebelumnya dengan mensitasi karya ilmiah yang lain. Dalam pembuatan kalimat sitasi, sering kali penulis menggunakan kalimat yang berorientasi objektif untuk dapat menghindari kritikan terhadap karya yang digunakan sebagai referensinya. Akan tetapi tidak menutup kemungkinan pengutipan dilakukan dengan orientasi positif yang bertujuan untuk mendukung karya ilmiah tersebut atau menggunakan orientasi negatif untuk melihat kekurangan dari karya ilmiah yang diacunya[1]. Sentimen di dalam kalimat sitasi juga dapat berguna untuk menentukan kualitas dari suatu karya ilmiah. Kualitas dari karya ilmiah seseorang dapat dinilai dengan melihat orientasi sitasi pada karya ilmiah yang merujuk karya ilmiah tersebut[2].

Untuk dapat memperlihatkan orientasi dari kalimat sitasi dapat digunakan analisis sentimen yang berguna untuk mengklasifikasikan orientasi dari sebuah kalimat berupa orientasi positif, negatif ataupun objektif. Pengelompokan ini menggunakan *classifier* untuk dapat mengklasifikasi orientasi yang terkandung didalam kalimat. Dikarenakan banyak teknik yang dapat digunakan untuk melakukan analisa sentimen[3], muncul sebuah permasalahan baru yaitu menentukan kombinasi ekstraksi fitur dan *classifier* yang dapat digunakan untuk menghasilkan model yang terbaik. Oleh karena itu dilakukanlah penelitian untuk mencari performansi setiap kombinasi antara *classifier* dan metode ekstraksi fitur. Pada pembuatan penelitian ini akan digunakan beberapa *classifier* yaitu *Support Vector Machine*(SVM), *Naïve Bayes* dan *Logistic Regresion*. Penggunaan *classifier* tersebut dikarenakan SVM, *Naïve Bayes*, *Logistic Regresion* ini memiliki performansi yang baik untuk mengklasifikasi sentimen[3-5]. SVM merupakan metode pembelajaran mesin yang bekerja atas *hyperplane* terbaik untuk memisahkan kelas kelas yang ingin didefinisikan. *hyperplane* ini sendiri merupakan garis yang memisahkan antara dua buah nilai[5]. Sedangkan *Naïve Bayes* merupakan sebuah teknik pembelajaran mesin berdasarkan probabilitas kemunculan kata didalam kalimat sitasi. Serta *Logistic Regresion* merupakan sebuah teknik pembelajaran mesin yang dapat memprediksi menggunakan *dependant variable*. Sedangkan ekstraksi fitur menggunakan *Word Level Features*, *Contextual Polarity Features*, *Sentence Structured Based Features*.

Penggunaan fitur-fitur tersebut dikarenakan didalam [1]. Didalam penelitian tersebut mereka melakukan kombinasi fitur 1-3 *grams* dengan fitur tambahan lainnya. Hasil penelitian tersebut didapat kombinasi yang terbaik adalah kombinasi fitur *unigram*, *bigram*, *trigram* + *Dependency parser* + *negation handling* dimana kombinasi tersebut memiliki *micro-f* sebesar 0.898. Didalam skenario penelitiannya fitur 1-3 *grams* dengan fitur tambahan lainnya, akan tetapi ketika performansi yang didapatkan oleh kombinasi fitur menjadi buruk, maka percobaan untuk kombinasi

fitur tersebut akan dihentikan dan dilanjut dengan kombinasi fitur lainnya. Hal tersebut mengakibatkan banyak kombinasi yang menjanjikan yang tidak diuji dan diketahui performansinya. Dengan menggunakan data dari kalimat sitasi yang terdapat didalam paper akan dilakukan penelitian terhadap kombinasi fitur dan *classifier* secara menyeluruh untuk kombinasi *unigram* dan *bigram* dan kombinasi *unigram*, *bigrams*, dan *trigrams*. Tujuan dari penelitian ini adalah untuk dapat menentukan kombinasi yang terbaik yang diukur berdasarkan performansi untuk setiap kombinasi, serta melakukan pengujian kombinasi fitur yang potensial dari [1] yang tidak diuji didalam penelitian tersebut.

1.2. Perumusan Masalah

Berdasarkan pada latar belakang diatas, rumusan masalah yang akan diuraikan dan diteliti adalah sebagai berikut:

1. Bagaimana pengaruh penggunaan kombinasi fitur terhadap performansi yang dihasilkan oleh *classifier* dalam klasifikasi sentimen pada kalimat sitasi.
2. Bagaimana pengaruh penggunaan *classifier* terhadap kasus analisis sentimen terhadap kalimat sitasi.

Dalam penelitian Tugas Akhir ini, terdapat beberapa batasan masalah dalam pembahasan dan pengerjaannya, diantaranya :

1. Kalimat sitasi dalam penelitian ini menggunakan dataset yang diperoleh dari [1] yang telah diekstraksi dan dilabel secara manual sebanyak 8736 kalimat dari 190 *paper* penelitian.
2. Menggunakan SVM, *Logistic Regression* dan *Naïve Bayes* sebagai *classifier* didalam penelitian ini
3. Menggunakan *Word Level Feature* berupa *1-3grams*, *Contextual Polarity Feature* dan *Sentence Structured Base Feature* berupa *Dependency Structures* serta *Negation Handling* sebagai ekstraksi fitur dalam penelitian ini
4. Menggunakan *SentiWordNet* dalam penanganan negasi.
5. Menggunakan *Polar-phrase* yang telah di ekstrak manual oleh Awais Athar dalam pelabelan *Contextual Polarity Feature*.

1.3. Tujuan

Terkait permasalahan yang dipaparkan diatas, dapat ditarik tujuan yang akan dicapai dalam Tugas Akhir ini, yaitu:

1. Mengetahui pengaruh penggunaan kombinasi fitur terhadap performansi yang dihasilkan oleh *classifier* dalam kasus analisa sentimen pada kalimat sitasi.
2. Mengetahui pengaruh penggunaan *classifier* terhadap kasus analisis sentimen pada kalimat sitasi.

1.4. Sistematika Penulisan

Metodologi yang digunakan dalam penyelesaian Tugas Akhir ini :

1. Studi Literatur

Pada tahap ini akan dilakukan pengumpulan informasi dari berbagai referensi baik dalam bentuk buku materi, ataupun jurnal ilmiah yang telah dipublikasikan baik yang berbahasa Indonesia maupun yang berbahasa Inggris.

2. *Data Preprocessing*

Pada tahap ini dataset yang terdiri dari 8736 kalimat yang telah di ekstrak dari 190 *paper* penelitian[2] akan dilakukan *data preprocessing* berupa menghilangkan tanda baca dan angka serta disimpan sesuai dengan input yang diminta oleh *tools* yaitu berupa *file CSV*.

3. Ekstraksi fitur dan Pengklasifikasian

Pada tahap ini akan mengimplementasikan metode Word Level Features, Contextual Polarity Features, Sentence Structured Based Features, dan Negation terhadap data yang sudah di Preprocessing serta melakukan pengkombinasi fitur yang telah didapatkan. Lalu melakukan pengklasifikasian terhadap hasil ekstraksi fitur menggunakan SVM, Naïve Bayes dan Logistic Regresion.

4. Analisa Hasil Pengujian

Pada tahap ini akan dianalisa hasil dari pengujian dengan menghitung F-Measure untuk setiap kombinasi sehingga untuk dapat dilihat akurasi dan ketepatan prediksi dari setiap kombinasi

5. Penarikan kesimpulan dan penyusunan laporan

Pada tahap ini akan dilakukan penarikan kesimpulan terhadap analisis yang telah dilakukan serta pembuatan laporan Tugas Akhir untuk mendokumentasikan setiap tahap tahap pengerjaan penelitian ini.