

## 1. Pendahuluan

### Latar Belakang

Pekembangan teknologi yang sangat cepat berpengaruh terhadap pertumbuhan dan penyebaran informasi. Teknologi internet mempunyai peranan penting dalam penyebaran informasi, salah satunya dalam menyampaikan opini terhadap suatu *movie*. Opini yang disampaikan akan menjadi sebuah informasi tentang *movie review*. Pengguna internet dapat mudah menemukan informasi berupa teks yang berkaitan dengan *movie* sebagai sarana acuan film apa yang akan ditonton. Dalam waktu satu bulan 69% penikmat *movie* bisa menonton satu *movie* dan 17,5% penikmat *movie* bisa menonton dua kali [6]. Semakin banyaknya informasi *movie review* yang tersebar di internet. Akibatnya banyak penikmat *movie* yang kesulitan menemukan informasi yang sesuai kebutuhannya. Informasi yang tersebar cenderung heterogen karena setiap orang memiliki pandangan yang subjektif dalam memahami suatu *movie*.

Berdasarkan permasalahan yang dihadapi oleh para penikmat *movie* maka penelitian yang dilakukan adalah analisis sentimen. Analisis sentimen adalah jenis pemrosesan bahasa alami untuk melacak *mood* publik tentang produk atau topik tertentu [10]. Tujuannya untuk menentukan polaritas dalam suatu *movie review* berdasarkan data yang membicarakan tentang suatu *movie*. Teknik analisis sentimen digunakan untuk mengekstrak dan mengelola data kemudian memasukkannya ke dalam kategori negatif atau positif. Analisis sentimen termasuk golongan *supervised learning*. Data *movie review* yang sudah dikumpulkan dijadikan data *training* dan data *testing*. Selanjutnya dilakukan proses klasifikasi.

Dalam penelitian ini klasifikasi yang digunakan adalah metode klasifikasi KNN (*K-Nearest Neighbor*). Metode KNN (*K-Nearest Neighbor*) yang digunakan pada sistem yang dibangun pada penelitian ini dipilih karena pada proses pencarian hasil klasifikasinya yang cukup mudah dan sederhana. Proses adaptasi dan pembelajaran pada klasifikasi KNN (*K-Nearest Neighbor*) menghasilkan nilai evaluasi yang tinggi [5]. KNN (*K-Nearest Neighbor*) adalah metode klasifikasi yang datanya ditentukan oleh jarak yang paling dekat dari objek tersebut. Kemudian ditambahkan metode *negation handling* untuk memperjelas polaritas kata yang terdapat dalam data *movie review*. Metode *negation handling* dipilih karena untuk menangani kata negasi yang kurang tepat dengan label datasetnya. Dalam penelitian sebelumnya tidak ada yang menggunakan *negation handling* pada sistem yang dibangun berdasarkan klasifikasi KNN (*K-Nearest Neighbor*) [8]. Ekstraksi fitur menggunakan metode TF-IDF untuk pembobotan hubungan kata. Pada penelitian lain yang telah dilakukan, metode TF-IDF digunakan pada analisis sentimen ulasan penjualan *marketplace* di Indonesia [3]. Kemudian untuk seleksi fitur menggunakan metode MI (*Mutual Information*). Dalam penelitian sebelumnya metode MI (*Mutual Information*) digunakan pada permasalahan klasifikasi dokumen untuk meningkatkan performansi dari sistem yang dibangun [7].

### Topik dan Batasannya

Dalam Penelitian ini membahas tentang penggunaan metode *negation handling* untuk mengatasi kata negasi yang terdapat di dataset *movie review*. Metode *negation handling* digunakan terhadap sistem yang dibangun berdasarkan klasifikasi KNN (*K-Nearest Neighbor*) untuk analisis sentimen dataset *movie review*. Dataset mempunyai dua tipe kategori yaitu sentimen positif dan sentimen negatif. Tahapan yang dilakukan yaitu *preprocessing* menggunakan metode *cleaning*, *case folding*, *tokenizing*, *stopword*, dan *stemming*. Menggunakan metode TF-IDF untuk mengekstraksi fitur dari dataset *movie review*. Metode MI (*Mutual Information*) digunakan sebagai seleksi fitur dataset yang sudah melalui tahap *negation handling*, *preprocessing*, dan ekstraksi fitur. Dataset berjumlah 7000 data berbahasa inggris yang terdiri dari 3500 *review* positif dan 3500 *review* negatif.

Batasan yang terdapat dalam penelitian ini adalah dataset yang digunakan terbatas dan tidak diambil semuanya dikarenakan keterbatasan perangkat keras pada penelitian ini. Dataset yang digunakan terbatas hanya berbahasa inggris.

### Tujuan

Penelitian yang dilakukan bertujuan untuk mengetahui performansi sistem yang dibangun dan pengaruh metode *negation handling* terhadap sistem yang dibangun menggunakan dataset *movie review*. Kemudian pengaruh metode *preprocessing stemming* pada kinerja sistem yang dibangun. Selanjutnya mengetahui performansi sistem yang didalamnya menggunakan metode MI (*Mutual Information*).

### Organisasi Tulisan

Bagian selanjutnya dalam penelitian ini adalah studi yang terkait dengan topik yang dibahas. Selanjutnya memaparkan mengenai sistem yang dibangun. Kemudian dilakukan analisis dari hasil pengujian dan melakukan evaluasi. Bagian terakhir adalah menarik kesimpulan yang didapat pada penelitian ini berdasarkan pengujian yang sudah dilakukan.