

## 1 Pendahuluan

### 1.1 Latar Belakang

Saat ini dikarenakan adanya Covid-19 dan *lockdown* yang mengakibatkan aktifitas sosial terbatas dan mencari hiburan diluar menjadi sulit didapat oleh pihak berwajib untuk menghindari penularan yang meluas. Alternatif hiburan yang tidak perlu keluar dari rumah yaitu adalah bermain *game* [1]. Steam adalah sebuah platform distributor game digital yang menyediakan platform untuk para '*Game Developer*' dari perusahaan besar (*Triple-A*) ataupun perusahaan kecil (*Indie*) untuk menjajakan *game*-nya sehingga bisa ditemukan oleh orang diseluruh dunia [2]. Steam juga memberikan tempat ulasan para user tentang *game* tersebut untuk memberikan pendapat mereka kepada pengembang *game*-nya sehingga dimungkinkan dapat meningkatkan kembali *game*-nya tetapi jumlah ulasan yang diberikan oleh user sangat banyak sehingga pengembang sangat sulit untuk mengetahui apakah user menyukai atau tidak menyukai *game* yang mereka buat dan juga beberapa dari ulasan tersebut tidak sesuai isi maknanya dengan yang mereka tulis, maka dari itu analisis sentimen digunakan untuk mempermudah pengembang dalam melihat ulasan yang telah diberikan user [3]. Dalam penelitian ini digunakan algoritma dan teknik menggunakan machine learning dalam analisis sentimen untuk menganalisa lebih dalam dari perbandingan input antara *Bigram* (dua kata) dan *Trigrams* (tiga kata) yang menggunakan berbagai algoritma, *machine learning* biasa digunakan untuk mengklasifikasi dan memprediksi dari sentimen baik itu positif atau negatif [4].

Tipe *machine learning* yang digunakan adalah *Supervised* yang menggunakan *training* dataset untuk membantu mengklasifikasi input data yang akan diuji. Berdasarkan dalam penelitian zuo dalam *paper* "*Sentiment Analysis of Steam Review Datasets using Naive Bayes and Decision Tree Classifier*" dari data yang dikumpulkan sebagian besar ulasan yang didapatkan pada steam menunjukkan *feedback* positif sedangkan sebagian kecilnya menunjukkan *feedback* negatif [5].

Pada penelitian ini akan dilakukan klasifikasi dengan menggunakan metode *Random forest* adapun metode ini dipilih berdasarkan penelitian yang telah dilakukan oleh Hartmann et.al [6] pada penelitiannya yang berjudul "*Comparing automated text classification methods*", penggunaan metode *Random Forest* menghasilkan performa klasifikasi yang lebih baik dari ANN, KNN, NB, dan SVM dengan akurasi mencapai 75%. Dan berdasarkan penelitian yang dilakukan oleh Ahuja et.al [7] Pada penelitiannya yang berjudul "*The Impact of Features Extraction on the Sentiment Analysis*" menunjukkan penggunaan TF-IDF pada pemerosesan analisis sentimen mendapatkan peningkatan hasil (3%-4%) lebih baik.

### 1.2 Topik dan Batasannya

Dalam penelitian analisis sentimen ini penulis mengambil dataset dari situs halaman Steam *review* dengan judul game *Cyberpunk 2077*. Sample yang diambil hanya 2000 data dan hanya diambil 1000 data yang terdiri dari 601 data positif dan 399 data negatif, dikarenakan waktu dan sumber daya yang terbatas dalam mengerjakan tugas akhir ini, dataset yang diambil hanya data ulasan yang berbahasa inggris karena rata-rata dari sample yang diambil berbahasa inggris.

### 1.3 Tujuan

Tujuan penelitian ini untuk melakukan analisa sentimen terhadap ulasan yang diberikan oleh user/pembeli dari game tersebut dan membandingkan pengaruh menggunakan Random forest dan fitur ekstraksi TF-IDF. Penelitian ini juga meneliti pengaruh menggunakan *Lemmatization* pada preprosesing terhadap hasil Analisis sentimen.

### 1.4 Organisasi Tulisan

Bagian selanjutnya pada penelitian ini adalah bagian 2 yang membahas Studi terkait dengan penelitian yang dilakukan, bagian 3 membahas tentang rancangan sistem yang dibangun, bagian 4 membahas hasil dari evaluasi yang diteliti, pada bagian 5 membahas kesimpulan dan saran untuk penelitian kedepannya.