

Pendahuluan

1.1 Latar Belakang

Umpan balik dari pengguna adalah tanggapan yang diberikan oleh pengguna untuk mencerminkan preferensi pada *recommender system*[1]. Bentuk umpan balik diantaranya komentar dan rating dari pengguna. Pada *e-commerce*, pengguna memberikan komentar dan rating untuk produk/item yang ditawarkan[2]. Salah satu manfaat dari umpan balik pengguna adalah untuk meningkatkan minat dan kepercayaan pelanggan lain [3].

Komentar merupakan opini dari user. Komentar bisa menjadi salah satu faktor yang mempengaruhi keputusan *user*[4]. Banyak *customer/user* yang membaca komentar secara acak dan akan membuat keputusan akan memakai produk/jasa tersebut atau tidak[5]. Semakin positif *review* atau komentar yang diberikan maka akan meningkatkan penjualan dan sebaliknya akan menurunkan penjualan [6].

Komentar dengan sentiment positif digambarkan dengan nilai rating yang tinggi dan sebaliknya komentar dengan sentiment negatif digambarkan rating rendah[7]. *User* memiliki kontrol bebas untuk memberikan rating sesuai atau tidak dengan sentiment komentar yang diberikan. Disisi lain, rating sesuai dengan rating digunakan sebagai alat ukur untuk mengukur reputasi produk pada *reputation system*[8]. Dengan demikian kesesuaian antara komentar dan rating perlu memiliki keselarasan nilai sehingga perlu dilakukan *mining* ulang.

Beberapa penelitian memanfaatkan *opinion mining* dari kata sebagai kumpulan kata (*corpus*) untuk memprediksi tingkat rating untuk suatu item berdasarkan komentar pada item tersebut[5][9]. *Opinion mining* atau *sentiment analysis* atau *review mining* merupakan suatu metode untuk mengevaluasi opini *user*, subjektif, emosi, *feedback* dan sebagainya yang diekspresikan dalam komentar[10]. *Sentiment analysis* memanfaatkan *deep neural network* untuk proses klasifikasi[11]

CNN adalah salah satu algoritma *deep neural network* yang populer[12]. Algoritma ini memiliki performansi yang sangat baik untuk permasalahan – permasalahan di *machine learning*, seperti *image classification*, *computer vision*, dan *natual language processing*(NLP)[12]. Beberapa penelitian sebelumnya memanfaatkan CNN untuk permasalahan NLP dan memperoleh hasil performansi yang baik[13]. *Review mining* merupakan salah satu permasalahan NLP. Dengan memanfaatkan CNN diharapkan dapat mendapatkan hasil performansi yang baik dalam penelitian ini.

1.2 Topik dan Batasannya

Berdasarkan latar belakang, komentar dan rating merupakan salah satu faktor yang berpengaruh dalam penentuan keputusan dari user. Kesesuaian sentiment komentar dengan nilai rating yang diberikan merupakan hal penting. Algoritma CNN diterapkan dalam memprediksi rating berdasarkan komentar. Tingkat performansi algoritma CNN akan diukur terhadap data set yang berbeda.

Batasan masalah dari penelitian ini yaitu melakukan prediksi dengan menggunakan algoritma CNN. Data set yang digunakan ada dua data set yaitu data set yelp dan data set amazon. Bahasa yang digunakan dalam data set yaitu bahasa inggris. Skala yang digunakan sebagai pengukuran rating yaitu 1,2,3,4, dan 5. Skala tersebut digunakan karena di dalam kedua data set menggunakan skala dengan jenis tersebut. Nilai skala 1 merupakan nilai terkecil dan 5 adalah nilai terbesar. Skala tersebut yang digunakan sebagai hasil output prediksi.

1.3 Tujuan

Adapun tujuan dari penelitian ini adalah untuk menerapkan algoritma CNN dalam proses prediksi rating dengan komentar. Selain itu, untuk mengukur performansi algoritma CNN yang diterapkan dalam memprediksi rating pada data set yelp dan amazon.

