

Analisis Klasifikasi Tweet Suatu Akun Film Production Dengan Kontent-based Dan Time-based Menggunakan Metode Naïve Bayes

Rizki Luthfan Azhari¹, Jondri², Kemas Muslim Lhaksamana³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹rizkiluthfan@students.telkomuniversity.ac.id, ²jondri@telkomuniversity.ac.id,

³kemasmuslim@telkomuniversity.ac.id

Abstrak

Pada era digital yang serba modern ini, media sosial menjadi sarana atau platform untuk menyebarkan berbagai macam informasi secara mudah. Twitter merupakan salah satunya, twitter sendiri adalah sebuah media sosial yang bisa menyebarkan suatu informasi melalui tweet (kata kata yang diunggah oleh pengguna). Tweet bisa mengandung berbagai macam informasi, pembahasan, video, gambar maupun tautan ke suatu website. Suatu tweet akan disebar dari satu pengguna ke pengguna lainnya dengan cara me-retweetnya. Pada penelitian ini bertujuan untuk menganalisa apakah suatu tweet akan di retweet oleh pengguna lainnya dengan menggunakan fitur kontent-based dan time-based dengan metode klasifikasi naïve bayes serta menggunakan k-fold cross validation dengan nilai k=5 untuk melakukan split data. Hasil performansi yang didapatkan dengan menerapkan metode tersebut berupa nilai rata-rata akurasi 61,36%, rata-rata precision yang didapatkan sebesar 65,06%, rata-rata untuk recall sebesar 55,61%, lalu rata-rata untuk f1-score sebesar 50,49%.

Abstract

In this modern digital era, social media has become a means or platform to easily disseminate various kinds of information. Twitter is one of them, twitter itself is a social media that can spread information through tweets (words uploaded by users). Tweets can contain various kinds of information, discussions, videos, images or links to a website. A tweet will be spread from one user to another by retweeting it. This study aims to analyze whether a tweet will be retweeted by other users using content-based and time-based features with the nave Bayes classification method and using k-fold cross validation with a value of k=5 to split the data. The performance results obtained by applying this method are in the form of an average accuracy value of 61.36%, an average precision obtained of 65.06%, an average for recall of 55.61%, then an average for f1-score. by 50.49%.

1. Pendahuluan

Latar Belakang

Twitter adalah sebuah *microblogging* dari Amerika dan sebuah media sosial yang dimana user bisa memposting dan berinteraksi satu sama lain menggunakan pesan bernama “*Tweets*”. Dalam *tweets* ini sendiri bisa berupa kata-kata atau teks, foto, video, maupun suatu tautan yang menghubungkan dengan media sosial lainnya.

Suatu perusahaan dalam dunia perfilman biasanya memiliki caranya tersendiri dalam memasarkan produk atau filmnya. Banyak *platform* yang bisa digunakan untuk melakukan advertising contohnya seperti iklan pada Youtube sesaat sebelum video yang dipilih oleh *user* dimulai, koran, majalah, radio dan sosial media (Instagram, Twitter, Facebook, etc). Pada sosial media biasanya perusahaan tersebut akan membuat suatu account yang akan memunculkan informasi mengenai film yang mereka produksi ke masyarakat luas. Informasi tersebut bisa berupa jadwal tayang, *trailer*, pemeran, dan sebagainya. *Account* yang mereka buat untuk pertama kali ini tidak ada bedanya dengan *account* biasa pada umumnya. Tapi seiring berjalannya waktu *account* tersebut akan memiliki banyak pengikut dikarenakan informasi yang mereka berikan lebih akurat dan lebih cepat jika dibandingkan dengan *account* lainnya. Oleh karena itu, Twitter *user* akan percaya bahwa *account* tersebut memang dikelola oleh perusahaan produsen film resmi, hal tersebut akan menghasilkan pengikut yang banyak untuk *account* perusahaan tersebut sehingga perusahaan tersebut akan bisa mengajukan ke Twitter untuk menjadi *account* resmi atau yang biasa dibilang dengan *official account*.

Difusi sendiri merupakan suatu proses yang dimana suatu informasi akan disebar atau diteruskan melalui suatu interaksi. Berdasarkan penelitian yang dilakukan oleh Liangjie Hong, Ovidiu Dan, dan Brian D. Davidson yang berjudul “*Predicting Popular Message in Twitter*”, informasi dari suatu *tweet* akan disebar ke user lain atau orang lain dengan cara me-*retweetnya*[1], sehingga informasi yang berada didalam *tweet* tersebut akan dengan mudah tersalurkan ke orang atau *user* lain. Pada penelitian[2] penulis mendapatkan hasil bahwa penelitiannya mendapatkan peningkatan F-measure sebesar 5% dibandingkan dengan *state of art* (secara statistik) dengan menggunakan 3 fitur yaitu keseluruhan dari berbasis pengguna, berbasis waktu, dan berbasis dan berbasis konten dengan menggunakan metode *random forest*. Lalu, pada penelitian[3] mengatakan bahwa metode *random forest*