# INTISARI

## KLASIFIKASI FILM BERDASARKAN KATA KUNCI ALUR MENGGUNAKAN *MULTI-LABEL K-NEAREST NEIGHBOR*

Oleh

Muhammad Amirul Mukminin
14/364230/PA/15958

Besarnya kuantitas film yang dirilis setiap tahun di seluruh dunia memberikan permasalahan baru dalam mengkategorikan film berdasarkan genrenya. Kategori ini bisa diklasifikasikan secara otomastis menggunakan metode yang terdapat dalam *data mining*. Dikarenakan suatu film bisa mengandung lebih dari satu genre, maka permasalahan klasifikasi film merupakan klasifikasi multilabel. Meskipun telah banyak penelitian sebelumnya yang bergantung pada berbagai atribur dalam film seperti: ringkasan alur, naskah, poster, dan cuplikan, penelitian ini mempertimbangkan penggunaan atribut lain yang belum pernah digunakan yaitu kata kunci alur. Model klasifikasi film dibuat menggunakan metode *Multilabel k-Nearest Neighbor* yang didasarkan atas kata kunci alur. Metode ini melatih model dengan cara menghitung probabilitas prior dan probabilitas posterior pada setiap genre untuk kemudian digunakan untuk memprediksi genre film yang belum diketahui genre-nya.

Dua parameter yang dipakai dalam model klasifikasi ini adalah nilai $k$ dalam *nearest neighbors (nn)* dan jenis perhitungan jarak dalam mengidentifikasi *nn*. Nilai $k$ berkisar antara 4 sampai 60 dan jenis perhitungan jarak yaitu cosine dan euclidean. Model klasifikasi yang dibuat berhasil memberikan nilai *f-measure* micro terbaik sebesar 67,16%. Nilai ini merupakan kisaran nilai terbaik yang bisa didapatkan oleh model klasifikasi genre film manapun dengan keluaran multilabel sampai saat ini. Didapat bahwa model menghasilkan kinerja terbaik ketika menggunakan nilai $k$ antara 20 sampai 30 dengan perhitungan jarak euclidean.

Kata-kata kunci: klasifikasi film, klasifikasi multilabel, *Multilabel k-Nearest Neighbor*, kata kunci alur.

# ABSTRACT

## MOVIE CLASSIFICATION BASED ON PLOT KEYWORDS USING MULTI-LABEL K-NEAREST NEIGHBOR

By

Muhammad Amirul Mukminin
14/364230/PA/15958

The large number of movies that released every year around the world provides a new problem for categorizing movies based on their genre. This category can be classified automatically by using the method in data mining. Because of a movie can contain more than one genre, so the problem of movie classification is a multilabel problem. Although there have been many studies used in various attribute films such as: plot summary, manuscript, poster and trailer, this study uses other attributes that have never been used that is plot keyword. The classification film model is made using Multilabel k-Nearest Neighbor method that use plot keyword as the data input. This method trains the model by calculating prior probabilities and posterior probabilities in each genre to predict the genre of unknown film.

Two parameters used in this model are the value of $k$ of the nearest neighbors (*nn*) and the type of distance calculation in identifying *nn*. The value of $k$ stands between 4 to 60 and the type of distance calculation is cosine and euclidean. The created classification model successfully gives the best f-measure micro value of 67.16%. This value is a highest number that can be obtained by genre film classification with output multilabel until now. It was found that the model produces the best performance when using a value of $k$ between 20 to 30 with the euclidean distance measurement.

Keywords: movie classification, multilabel classification, Multilabel k-Nearest Neighbor, plot keyword.