

# Prediksi Rating Drama Korea Menggunakan Algoritma Iterative Dichotomiser 3

Hidayatin Novi Nurfajriah\*, Wina Witanti, Rezki Yuniarti

Jurusan Informatika, Fakultas MIPA  
Universitas Jenderal Achmad Yani  
Jl. Terusan Sudirman, Cimahi  
hidayatin.hnn@gmail.com

**Abstrak**—Drama korea merupakan salah satu media penyebaran korean wave yang memiliki jumlah penggemar atau penonton yang banyak. Salah satu faktor yang menentukan dalam pemilihan drama yaitu rating. Rating merupakan nilai yang menggambarkan popularitas suatu program di media pada saat tertentu. Perhitungan rating dilakukan bersamaan dengan penayangan program tersebut. Prediksi merupakan proses memperkirakan sesuatu dengan mengoreksi aksi sebelumnya, untuk meramalkan suatu kondisi dibutuhkan perhitungan yang tepat guna menjawab permasalahan tersebut. Perhitungan tersebut dapat menggunakan Algoritma Iterative Dichotomiser 3 (ID3). Dengan memanfaatkan data historis rating drama yang telah tayang, maka rating drama baru yang akan tayang dapat diprediksi nilai ratingnya. Penelitian ini bertujuan untuk membangun sistem yang dapat memprediksi nilai rating yang dinyatakan dengan status rating (tinggi atau rendah) dari drama baru dengan menggunakan Algoritma Iterative Dichotomiser 3 (ID3).

**Kata kunci**—drama korea; rating; prediksi; algoritma iterative dichotomiser 3.

## I. PENDAHULUAN

Rating merupakan satuan ukuran yang digunakan untuk mengukur performa dari stasiun TV. Setiap stasiun TV berlomba menayangkan program-program yang dapat menarik minat pemirsanya agar mendapat performa rating yang lebih dibanding TV lain. Bagi tim produksi, rating menjadi salah satu tolak ukur apakah drama yang diproduksi digemari atau tidak. Perhitungan rating dilakukan untuk setiap episode yang tayang. Tinggi atau rendahnya rating dapat mempengaruhi proses produksi.

Drama korea sendiri mulai populer di seluruh Asia (termasuk Indonesia) dan memberi kontribusi pada fenomena *Korean Wave*. Pertama kali drama korea masuk ke Indonesia pada tahun 2002 dan mulai terkenal pada tahun 2009 lewat tayangan *Boys Before Flower* (BBF) yang ditayangkan di stasiun televisi Indosiar[1],[2]. Dengan perkembangan teknologi selain di stasiun televisi, drama korea dapat juga ditonton secara *online* di berbagai *website*. Tidak hanya dapat menonton drama saja namun terdapat berbagai informasi tentang drama korea, baik drama yang sudah tayang maupun drama yang akan tayang.

Perhitungan rating program televisi di Korea sekarang ini dilakukan oleh dua lembaga, yaitu AGB Nielsen dan TNmS. Proses perhitungan rating dilakukan bersamaan dengan penayangan program atau drama tersebut. Perhitungan rating dilakukan selama program ditayangkan, nilai rating akan diberikan kepada setiap stasiun TV untuk semua program yang tayang pada hari tersebut.

Selain untuk tim produksi, nilai rating juga sangat penting untuk penggemar drama korea yang sedang memilih drama mana yang akan ditontonnya. Prediksi mempertimbangkan nilai yang belum terlihat pada masa akan datang berdasarkan pola-pola yang terjadi sebelumnya. Prediksi merupakan proses memperkirakan sesuatu dengan mengoreksi aksi sebelumnya, untuk meramalkan suatu kondisi dibutuhkan perhitungan yang tepat guna menjawab permasalahan tersebut. Perhitungan tersebut dapat menggunakan Algoritma Iterative Dichotomiser 3 (ID3), dengan memanfaatkan data historis rating drama yang telah tayang, maka rating drama baru yang akan tayang dapat diprediksi nilai ratingnya.

Beberapa penelitian prediksi terhadap rating film yang akan diberikan *user* dengan menggunakan metode *Counterpropagation Network* memiliki akurasi tingkat kesalahan terbesar 0.875 untuk data *testing* dan 0.963 untuk data uji[3]. Penelitian lain tentang rating sudah dilakukan sebelumnya, penelitian tersebut menganalisa teks berisi ulasan film untuk diubah menjadi rating (dalam bentuk bintang). Hasil penilaian sistem sebesar 97% didapat dari hasil pengujian *black box testing*[4]. Penelitian Algoritma Iterative Dichotomiser 3 (ID3) untuk mengidentifikasi data rekam medis dengan kasus penyakit Diabetes Melitus memiliki akurasi 89,759% berdasarkan pengukuran kinerja klasifikasi dan 72,619% berdasarkan pengukuran akurasi hasil klasifikasi algoritma ID3 menggunakan data uji sebanyak 84 sampel[5]. Penelitian lain tentang ID3 sudah dilakukan dalam bidang pendidikan, penelitian ini memperlihatkan pemakaian pohon keputusan untuk mempermudah pengambilan keputusan penerimaan mahasiswa baru[6].

Pohon keputusan dalam Algoritma ID3 dipengaruhi oleh jumlah data latih dan variasi data yang digunakan, semakin banyak data latih yang digunakan maka proses pelatihan akan semakin lama dan jumlah data latih mempengaruhi tingkat akurasi. Data yang digunakan adalah data drama yang tayang mulai tahun 2010 sampai dengan Desember 2016.

Tujuan penelitian ini, yaitu untuk membangun sistem yang dapat memprediksi nilai rating dari drama baru yang akan tayang mendapatkan rating tinggi atau rendah dengan Algoritma ID3, sehingga dapat mempermudah penggemar drama korea memilih drama apa yang selanjutnya akan di tonton. Berdasarkan hal tersebut, maka pada penelitian ini dirancang sebuah sistem prediksi drama korea mendapatkan rating tinggi atau rendah dengan Algoritma ID3.

## II. METODE

Pada penelitian ini terdiri dari beberapa tahapan sistematis untuk menghasilkan prediksi rating drama baru yang dinyatakan dengan status rating tinggi atau rendah menggunakan Algoritma ID3. Tahapan tersebut antara lain mengidentifikasi data drama korea yang akan digunakan sebagai data latih.

### A. Pengumpulan Data

Pengumpulan data dilakukan dengan cara studi dokumentasi yaitu teknik pengumpulan data yang tidak langsung pada subjek penelitian, melainkan pada dokumen-dokumen tertentu. Jenis data yang digunakan dalam penelitian ini merupakan data sekunder. Data tersebut merupakan drama yang tayang mulai 2010 sampai dengan Desember 2016.

### B. Algoritma Iterative Dichotomiser 3 (ID3)

Algoritma Iterative Dichotomiser 3 (ID3) pertama kali diperkenalkan oleh Quinlan pada tahun 1986 yang digunakan untuk menginduksi *decision tree*. Algoritma ID3 dapat bekerja dengan baik pada semua fitur yang mempunyai tipe data kategorikal (nominal atau ordinal). Dalam perkembangannya, ID3 banyak mengalami perbaikan pada versi berikutnya seperti C4.5 atau C5.0.

Berikut adalah langkah-langkah dari Algoritma ID3:

1. Dimulai dari *node* akar.
2. Untuk semua fitur, hitung nilai *entropy* untuk semua sampel (data latih) pada *node*.
3. Pilih fitur dengan *information gain* yang maksimal.
4. Gunakan fitur tersebut sebagai *node* pemecahan menjadi cabang.
5. Lakukan secara rekursif pada setiap cabang yang dibuat dengan mengulangi langkah 2 sampai 4 hingga semua data dalam setiap *node* hanya memberikan satu label kelas. *Node* yang tidak dapat dipecah lagi merupakan *leaf* (daun) yang berisi keputusan (label kelas).

*Entropy* dapat dihitung menggunakan persamaan berikut:

$$E(s) = - \sum_{i=1}^m p(\omega_i|s) \log_2 p(\omega_i|s) \quad (1)$$

Di mana  $p(\omega_i|s)$  adalah proporsi kelas ke-I dalam semua data latih yang diproses di *node* s.  $p(\omega_i|s)$  didapat dari jumlah semua baris data dengan label kelas *i* dibagi jumlah baris semua data. Sementara *m* adalah jumlah nilai berbeda dalam data.

*Entropy* digunakan untuk menentukan *node* manakah yang akan menjadi pemecah data latih berikutnya. Nilai *entropy* yang lebih tinggi akan meningkatkan potensi klasifikasi. Apabila *entropy* untuk *node* bernilai 0 berarti semua data vektor berada pada label kelas yang sama dan *node* tersebut menjadi daun yang berisi keputusan (label kelas), yang perlu diperhatikan dalam perhitungan *entropy* adalah jika salah satu dari elemen  $\omega_i$  jumlahnya 0 maka *entropy* dipastikan bernilai 0. Jika proporsi semua elemen  $\omega_i$  sama jumlahnya maka dipastikan *entropy* bernilai 1.

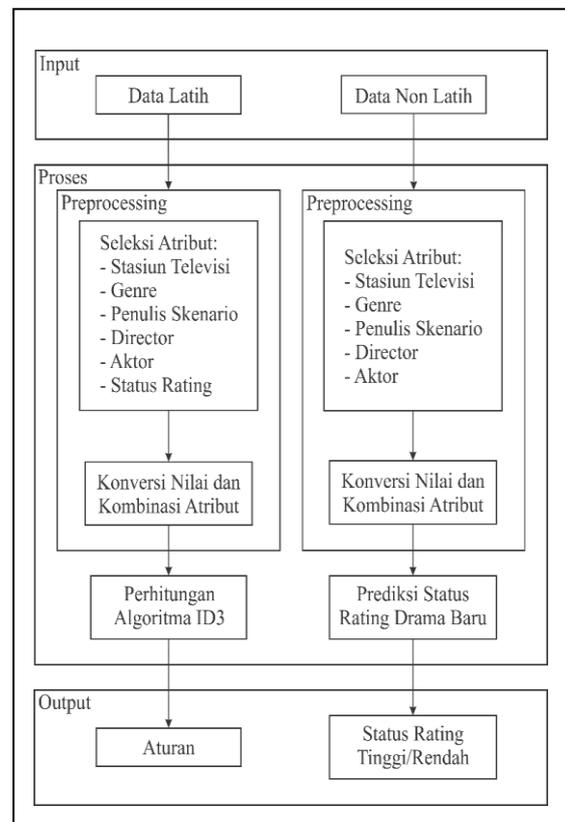
*Gain* digunakan untuk memperkirakan pemilihan fitur yang tepat untuk menjadi pemecah pada *node* tersebut. *Gain* sebuah fitur ke-*j* dihitung menggunakan persamaan 2:

$$G(s, j) = E(s) - \sum_{i=1}^n p(v_i|s) \times E(s_i) \quad (2)$$

$p(v_i|s)$  adalah proporsi nilai *v* yang muncul pada kelas *node*.  $E(s_i)$  adalah *entropy* komposisi nilai *v* dari kelas ke-*j* dalam data ke-*i* *node* tersebut. *n* adalah jumlah nilai berbeda dalam *node*.

### C. Perancangan Sistem

Pada sistem ini terdapat dua proses utama, yaitu proses pelatihan dan proses prediksi. Gambar 1 menunjukkan proses penyelesaian masalah menggunakan algoritma ID3.



Gambar 1. Gambaran umum proses prediksi

1) *Input*

*Input* pada penelitian ini adalah data drama yang tayang mulai tahun 2010 sampai Desember 2016 untuk data latih dan data drama yang tayang tahun 2017 (Januari s/d Desember). Data *input* tersebut didapat dari *website* asianwiki dan wikipedia. Tabel 1 merupakan deskripsi fitur yang digunakan untuk *input*.

TABEL 1. DESKRIPSI FITUR

No	Fitur	Nilai Fitur	Konversi Nilai Fitur
1.	Stasiun TV	JTBC	JTBC
		OCN	OCN
		TVN	TVN
		KBS	KBS
		SBS	SBS
		MBC	MBC
2.	Genre	Nilai Fitur Genre	Nama Genre
3.	Writer	Nilai Fitur Writer	ID Writer
4.	Director	Nilai Fitur Director	ID Director
5.	Actor	Nilai Fitur Actor	ID Actor
6.	Status Rating	- JTBC, OCN, dan TVN	
		< 6 %	Rendah
		>= 6 %	Tinggi
		- KBS, SBS, dan MBC	
		< 15 %	Rendah
	>= 15 %	Tinggi	

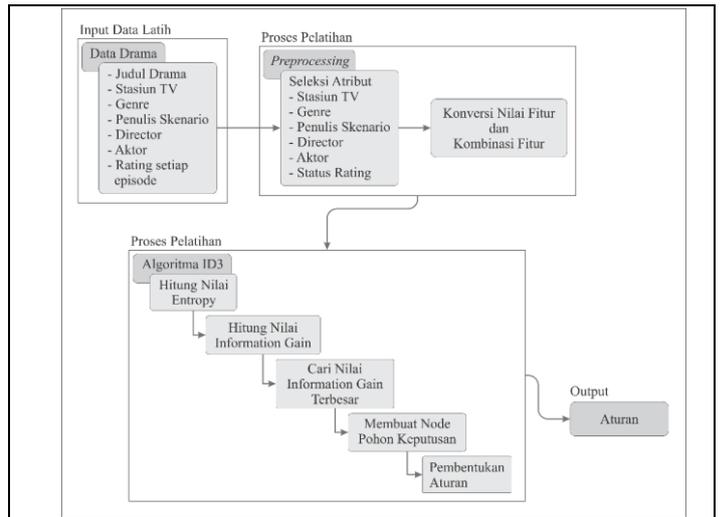
Ket: Nilai fitur *writer*, *director*, dan *actor* dikonversi menjadi id dari nilai fitur tersebut.

2) *Proses*

Terdapat dua proses yang dilakukan dalam penelitian ini, yaitu proses yang dilalui data latih dan data non latih. Pada proses pelatihan data, *input* merupakan data drama yang tayang mulai tahun 2010 sampai Desember 2016. Proses pertama dilakukan pada proses pelatihan adalah *preprocessing*. Pada tahap *preprocessing* dilakukan proses seleksi atribut, kemudian data dikonversi dan dikombinasi nilai fiturnya. Seleksi atribut merupakan proses pengambilan data yang akan digunakan. Proses konversi nilai fitur merupakan proses penyesuaian nilai rating drama menjadi tinggi atau rendah sesuai asumsi yang telah ditentukan. Proses kombinasi data dilakukan untuk menyesuaikan keseluruhan isi dari nilai fitur pada setiap judul drama. Proses ini diperlukan karena adanya nilai fitur yang lebih dari satu pada setiap judul drama.

Proses selanjutnya yang dilakukan pada proses pelatihan yaitu proses perhitungan algoritma ID3. Proses ini memiliki beberapa sub proses diantaranya, menghitung nilai *entropy* untuk semua sampel, menghitung nilai *information gain* setiap fitur, mencari nilai *information gain* terbesar untuk dijadikan

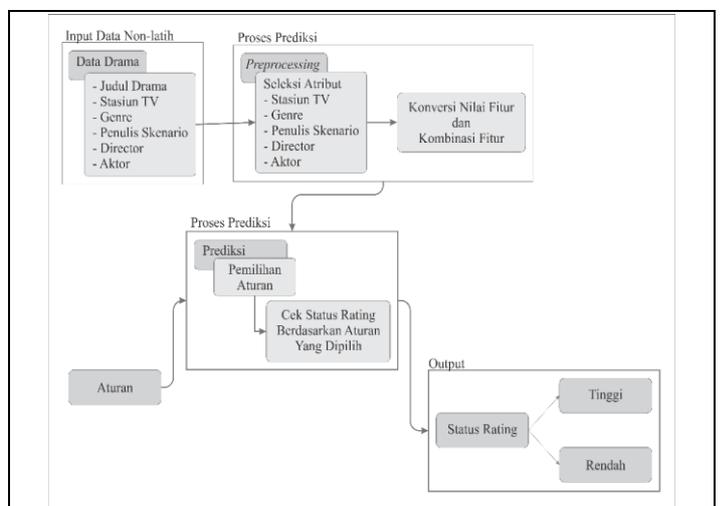
*node* pemecah (*node* akar dalam pohon keputusan), membentuk pohon keputusan yang akan menghasilkan *output* berupa aturan. Nilai *entropy* dan *information gain* dihitung dengan menggunakan rumus pada persamaan (1) dan (2). Gambar 2 menjelaskan gambaran umum proses pelatihan.



Gambar 2. Gambaran umum proses pelatihan

Proses prediksi menggunakan data *input* drama korea yang tayang dari Januari s/d Desember 2017. Data *input* ini melewati proses yang sama dengan data *input* yang digunakan pada proses pelatihan, yaitu melalui tahap *preprocessing*. Tahap *preprocessing* yang dilalui oleh data non latih atau data prediksi adalah pengambilan data atribut drama yang dimasukkan oleh pengguna, di mana data atribut drama berupa stasiun TV, *writer*, *director*, dan *actor*.

Proses setelah *preprocessing* yaitu konversi nilai dan kombinasi atribut. Hasil dari tahap *preprocessing* digunakan dalam proses prediksi. Pada proses prediksi dilakukan pemilihan aturan dengan menggunakan aturan yang dihasilkan pada proses pelatihan. Keluaran dari proses prediksi berupa status rating yang dinyatakan dengan rendah atau tinggi. Gambaran umum proses prediksi diuraikan pada Gambar 3.



Gambar 3. Gambaran umum proses prediksi

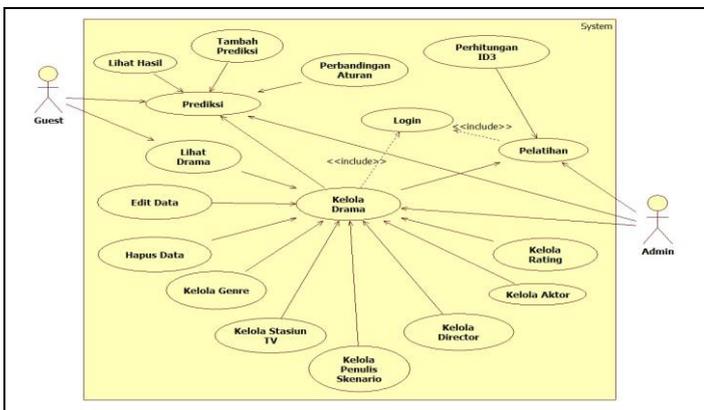
### 1) Output

Output yaitu prediksi status rating drama baru. Hasil prediksi yaitu tinggi dan rendah.

Terdapat dua proses utama dalam penelitian ini yaitu proses yang dilewati data latih dan proses yang dilewati data non latih. Pada proses yang dilewati data latih terdapat dua sub proses yaitu tahap *preprocessing* dan tahap perhitungan Algoritma ID-3. Sedangkan proses yang dilewati data non latih adalah tahap *preprocessing* dan tahap prediksi status rating drama baru.

### D. Use Case Diagram

Use Case diagram merupakan gambaran fungsional perangkat lunak atau layanan yang akan dirancang antara pengguna dengan perangkat lunak. Dalam perancangannya *use case* diagram membutuhkan aktor. Pada penelitian ini terdapat dua aktor yang berperan dalam sistem, yaitu *guest* dan *admin*. *Guest* merupakan penonton atau penggemar drama korea, dalam sistem ini *guest* dapat melihat drama dan atributnya, serta melakukan prediksi dan melihat hasil prediksi. *Admin* merupakan pihak pengguna yang memiliki hak akses penuh dalam perangkat lunak. *Admin* dapat mengelola semua *use case* yang ada. Gambaran *use case* yang dibangun dapat dilihat pada Gambar 4.



Gambar 4. Use Case Diagram Sistem Prediksi Rating

## III. HASIL DAN DISKUSI

### A. Pembahasan

Implementasi algoritma ID3 untuk memprediksi rating drama korea dilakukan melalui beberapa tahap seperti yang telah dibahas pada bagian perancangan sistem. Dataset yang digunakan adalah data drama yang tayang mulai tahun 2010 sampai dengan Desember 2016. Data *input* tersebut memiliki fitur stasiun TV dengan 6 nilai fitur, *genre* dengan 42 nilai fitur, penulis *scenario* (*writer*) memiliki 211 nilai fitur, director dengan 189 nilai fitur, aktor dengan 1778 nilai fitur, dan status rating dengan dua nilai fitur (tinggi dan rendah). Data tersebut diolah untuk menghasilkan data kombinasi, yang kemudian akan dikonversi dengan sesuai aturan yang telah ditetapkan. Data hasil konversi digunakan untuk *input* data latih yang akan diolah pada proses algoritma ID3. Data latih yang digunakan pada sistem ini adalah 249 judul drama, yang menghasilkan 21077 kombinasi data latih.

Proses algoritma ID3 dimulai dengan menghitung nilai *entropy* dan *information gain*. Dari perhitungan tersebut diketahui bahwa fitur yang paling berpengaruh terhadap rating adalah *Writer*. Hal ini ditunjukkan dengan nilai *information gain* yang didapat sebesar 0.593. Atribut tersebut digunakan sebagai *node* pemecah (*node* akar dalam pohon keputusan), yang nantinya akan diubah menjadi aturan. Proses perhitungan nilai *entropy* dan nilai *information gain* diulang dengan patokan nilai *information gain* terbesar sebelumnya. Pada perhitungan selanjutnya fitur yang memiliki nilai *information gain* terbesar tidak dilibatkan dalam perhitungan.

### B. Skenario Pengujian

Pengujian yang akan dilakukan pada penelitian ini adalah dengan melakukan penghitungan *precision*, *recall*, *f-measure*, dan *accuracy*. Pengujian akan dilakukan sebanyak 8 kali dengan perbandingan antara data latih dan data uji yang beragam.

### C. Data Pengujian

Data drama yang digunakan diambil dari [www.wikipedia.com](http://www.wikipedia.com) dan [www.asianwiki.com](http://www.asianwiki.com). Data yang digunakan merupakan drama yang tayang dari Januari sampai dengan Desember 2017 untuk data uji, sedangkan untuk data latih digunakan drama yang tayang dari tahun 2010 sampai Desember 2016. Data latih yang digunakan untuk pengujian sebanyak 200 judul drama dan 10 judul drama untuk data uji.

### D. Lingkungan Pengujian

Pengujian akan dilakukan sebanyak 8 kali. Jumlah data uji yang akan digunakan dibuat tetap. Hal ini bertujuan untuk melihat pengaruh jumlah data latih terhadap hasil dan evaluasinya perbandingan data latih dan data uji, serta jumlah data yang digunakan dapat dilihat pada TABEL 2.

TABEL 2. TABEL DATASET PENGUJIAN

No	Perbandingan Jumlah Data		Jumlah Data Sebenarnya	
	Data Latih	Data Uji	Data Latih	Data Uji
1.	1	1	10	10
2.	2	1	20	10
3.	4	1	40	10
4.	6	1	60	10
5.	8	1	80	10
6.	10	1	100	10
7.	15	1	150	10
8.	20	1	200	10

## IV. KESIMPULAN

Penelitian ini telah menghasilkan sebuah sistem prediksi rating drama korea. Sistem ini dapat memprediksi status rating drama baru yang dapat digunakan sebagai rujukan untuk penggemar drama korea yang akan memilih drama baru mana yang akan ditonton. Proses pada sistem ini menggunakan algoritma ID3 untuk membentuk pohon aturan yang menggambarkan pola atribut yang mempengaruhi status rating. Sistem ini menghasilkan *output* status rating drama yang dinyatakan dengan Tinggi dan Rendah.

## DAFTAR PUSTAKA

- [1] Y. Noor Rahmah, *Diplomasi Kebudayaan Republic of Korea Melalui Film dan Drama : Pencapaian Kepentingan Citra dan Ekonomi Republic of Korea di Indonesia*, Jakarta: Universitas Islam Negeri Syarif Hidayatullah, 2013.
- [2] A. Nesya, *Kebudayaan Populer Korea : Hallyu dan Perkembangannya di Indonesia*, Depok: Universitas Indonesia, 2010.
- [3] H. Yoga, D. Retno Novi, and B. Moch Arif, *Prediksi Rating Film Menggunakan Metode Counterpropagation Network*, Bandung: Universitas Telkom, 2008.
- [4] U. Bobby Chahya, H. Tacbir, and R. Faiza, "Penerapan Knowledge Management System untuk Penetapan Rating Film Layar Lebar Menggunakan Algoritma TF-IDF". Skripsi, Sarjana Informatika. Fakultas MIPA. Universitas Jenderal Achmad Yani. Cimahi. Unpublished.
- [5] T. Avia Enggar, I. Dwi, and H. Abdul, "Algoritma Iterative Dichotomiser 3 (ID3) Untuk Mengidentifikasi Data Rekam Medis," *Jurnal Gaussian*, vol. 4 Nomor 2, pp. 237-246, 2015.
- [6] Wahyudin, "Metode Iterative Dichotomiser 3 (ID3) untuk Penyeleksian Mahasiswa Baru," *Jurnal Pendidikan Teknologi Informasi dan Komunikasi (PTIK)*, vol. 2 No.02, Desember, pp. 5-15, 2009.
- [7] C. Vinita, W. Romi Satria, and Purwanto, "Komparasi Algoritma Klasifikasi Machine Learning dan Feature Selection pada Analisis Sentimen Review Film," *Journal of Intelligent System*, vol. 1 No. 1, February, pp. 56-60, 2015.
- [8] S. Juice Adiana, S. Eko, and N. Oky Dwi, "Pendekatan Metode Pohon Keputusan Menggunakan Algoritma ID3 untuk Sistem Informasi Pengukuran Kinerja PNS," *Jurnal Sistem Informasi Bisnis*, vol. 02, pp. 75-86, 2014.
- [9] K. Obbie, *Penerapan Algoritma Klasifikasi Data Mining ID3 untuk Menentukan Penjurusan Siswa SMAN 6 Semarang*, Semarang.
- [10] W. Budiwan, M. Lailil, and R. Achmad, *Klasifikasi Jurnal Ilmiah Berbahasa Inggris Berdasarkan Abstrak Menggunakan Algoritma ID3*, Malang, 2010. Unpublished.
- [11] D. Ressi, "Korean Wave, Imperialisme Budaya, dan Komersialisasi Media," *Jurnal Komunikasi Universitas Tarumanagara*, vol. VI, no. 01, pp. 41-51, 2014.