

ABSTRAK

Perkembangan industri *e-sport* yang ada di Indonesia telah mendapatkan respon positif dari pemerintah. IeSPA sebagai wadah untuk menaungi komunitas dan para pemain agar dapat mengembangkan bakat dan kemampuan dalam dunia game yang kompetitif. Dunia *e-sport* cukup menarik banyak perhatian masyarakat khususnya bagi para pecinta video game. Hal ini menimbulkan opini-opini dari masyarakat tentang dunia *e-sport* yang telah menjadi trend baru masa kini. Twitter menjadi sarana untuk mempublikasikan opini tersebut dalam hal ini berbentuk tweet. Penelitian ini menawarkan pengumpulan sentimen atau opini dari masyarakat mengenai dunia *e-sport* di Indonesia. Data dikumpulkan dari media sosial Twitter berupa tweet. Data lalu diolah dan diberikan nilai di setiap kata. Pembobotan kata dilakukan dengan menggunakan metode TF-IDF yang berperan sebagai ekstraksi fitur. Pengklasifikasian dilakukan dengan menggunakan *Naïve Bayes* serta menggunakan *K-Fold Cross Validation* sebagai metode untuk melakukan evaluasi model. Output yang dihasilkan nantinya bisa menjadi tolak ukur dalam pengambilan keputusan bagi instansi terkait maupun masyarakat. Berdasarkan hasil penelitian tersebut dapat diketahui bahwa sentimen masyarakat cenderung positif dengan rata-rata akurasi yang didapatkan sebesar 74.76%. Dapat disimpulkan sentimen masyarakat tentang dunia *e-sport* cukup baik.

Kata kunci: Sentimen analisis, *E-sport*, *Naive Bayes*, TF-IDF.

DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN ORISINALITAS.....	iv
ABSTRAK.....	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR KODE PROGRAM.....	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.2.1 Permasalahan.....	3
1.2.2 Usulan Solusi	3
1.2.3 Pertanyaan Penelitian	4
1.3 Tujuan dan Manfaat.....	4
1.3.1 Tujuan	4
1.3.2 Manfaat	4
1.4 Batasan Masalah.....	4
1.5 Sistematika Proposal	5
BAB II KAJIAN PUSTAKA	6
2.1 Analisis Sentimen.....	6
2.2 Twitter	7
2.3 <i>E-sport</i>	7
2.4 <i>Data Crawling</i>	8
2.5 <i>Text Preprocessing</i>	9
2.6 Pembobotan TF-IDF.....	9
2.7 <i>Naïve Bayes</i>	10
2.8 Evaluasi Model.....	11
2.9 Penelitian Terkait	14
BAB III METODOLOGI PENELITIAN.....	17

3.1	Rancangan Sistem	17
3.1.1	Data <i>Crawling</i> Berdasarkan <i>Keyword</i>	18
3.1.2	Pelabelan Pada Data	18
3.1.3	Data Preprocessing	19
3.1.4	Pembobotan TF-IDF	21
3.1.5	Pembagian Data	23
3.1.6	Klasifikasi <i>Naive Bayes</i>	24
3.1.7	Evaluasi Model.....	25
3.2	Skenario Uji Coba	26
3.2.1	Skenario I	26
3.2.2	Skenario II.....	27
BAB IV HASIL DAN PEMBAHASAN		28
4.1	Lingkungan Uji Coba	28
4.2	Hasil Pengumpulan Data	28
4.3	Tahapan <i>Preprocessing</i>	29
4.3.1	<i>Remove Duplicates</i>	30
4.3.2	<i>Cleaning</i>	31
4.3.3	<i>Tokenizing</i>	32
4.3.4	<i>Transform Case</i>	33
4.3.5	<i>Filter Token (by Length)</i>	34
4.3.6	<i>Stemming</i>	35
4.3.7	<i>Stopword Removal</i>	36
4.4	Tahapan Pembobotan Kata.....	38
4.5	Tahapan Klasifikasi	39
4.6	Tahapan Evaluasi	42
4.7	Tampilan Sistem.....	43
4.7.1	Tampilan Tab <i>Dataset</i>	43
4.7.2	Tampilan Tab <i>Preprocessing</i>	44
4.7.3	Tampilan Tab Pembobotan Kata.....	45
4.7.4	Tampilan Tab Klasifikasi.....	45
4.8	Analisis Skenario Uji Coba	46
4.8.1	Hasil Uji Coba Skenario I	46

4.8.2 Hasil Uji Coba Skenario II.....	49
BAB V PENUTUP.....	52
5.1 Kesimpulan.....	52
5.2 Saran.....	52
DAFTAR PUSTAKA	53

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi K-Fold Cross Validation	12
Gambar 2.2 Contoh Confusion Matrix	13
Gambar 3.1 Diagram Alur Rancangan Sistem	17
Gambar 3.2 Flowchart Preprocessing	19
Gambar 3.3 Flowchart TF-IDF	21
Gambar 3.4 Ilustrasi K-Fold Cross Validation	23
Gambar 3.5 Flowchart Naive Bayes	24
Gambar 4.1 Hasil Remove Duplicates	30
Gambar 4.2 Hasil Cleaning	32
Gambar 4.3 Hasil Tokenizing	33
Gambar 4.4 Hasil Transform Case	34
Gambar 4.5 Hasil Filter Token	35
Gambar 4.6 Hasil Stemming	36
Gambar 4.7 Hasil Stopword Removal	38
Gambar 4.8 Jumlah Kata Sebelum dan Sesudah Preprocessing	38
Gambar 4.9 Hasil Pembobotan Kata	39
Gambar 4.10 Hasil Pembobotan Kata	41
Gambar 4.11 Tampilan Tab Dataset	44
Gambar 4.12 Tampilan Tab Preprocessing	44
Gambar 4.13 Tampilan Tab Pembobotan Kata	45
Gambar 4.14 Tampilan Tab Klasifikasi	46
Gambar 4.15 Hasil Uji Coba Skenario 1	47
Gambar 4.16 Proses Uji Coba Skenario II	49
Gambar 4.17 Hasil Uji Coba Skenario II	50

DAFTAR TABEL

Tabel 2.1 Penelitian Terkait	15
Tabel 3.1 Contoh Data Tweet Setelah Pelabelan	18
Tabel 3.2 Contoh data hasil preprocessing	22
Tabel 3.3 Contoh data hasil pembobotan TF-IDF	22
Tabel 3.4 Contoh Confusion Matrix	25
Tabel 3.5 Contoh nilai TP, FP, TN, dan FN tiap kelas	25
Tabel 3.6 Contoh hasil penerapan pada Skenario I.....	26
Tabel 3.7 Contoh hyperparameter tuning pada Skenario II	27
Tabel 4.1 Spesifikasi Perangkat Keras dan Perangkat Lunak.....	28
Tabel 4.2 Jumlah data tweet berdasarkan kata kunci yang digunakan	29
Tabel 4.3 Contoh data yang didapatkan.....	29
Tabel 4.4 Detail Hasil Uji Coba Skenario 1.....	47

DAFTAR KODE PROGRAM

Kode Program 4.1 Menghitung Jumlah Kata.....	30
Kode Program 4.2 Remove Duplicates.....	30
Kode Program 4.3 Cleaning.....	32
Kode Program 4.4 Tokenizing	32
Kode Program 4.5 Transform Case.....	33
Kode Program 4.6 Filter Token	34
Kode Program 4.7 Stemming.....	36
Kode Program 4.8 Stopword Removal	37
Kode Program 4.9 Pembobotan Kata TF-IDF	39
Kode Program 4.10 Proses Klasifikasi	41
Kode Program 4.11 Proses Evaluasi	43