

ABSTRAK

Kemajuan teknologi telah berdampak signifikan pada kehidupan manusia, termasuk dalam sektor transportasi *online*. Layanan transportasi online seperti Grab, Gojek, InDrive, dan Maxim telah menjadi populer di Indonesia. Dengan Grab sebagai salah satu aplikasi yang sangat banyak digunakan di Asia Tenggara, pengguna aplikasi Grab sering memberikan ulasan melalui *Google Play Store* dan ulasan ini menjadi sumber informasi penting bagi calon pengguna dalam menilai kualitas aplikasi. Penelitian ini bertujuan untuk menganalisis ulasan pengguna terhadap aplikasi Grab di *Google Play Store*. Analisis dilakukan menggunakan 1.658 data, di antaranya 512 ulasan positif dan 1.146 ulasan negatif. Penelitian ini diawali dengan proses *Scraping* data, seleksi data, pelabelan data, *Text Preprocessing*, ekstraksi fitur *TF-IDF*, penyeimbangan data menggunakan *SMOTE*, seleksi fitur menggunakan *Chi Square*, pembagian data dan proses klasifikasi menggunakan metode *Random Forest* kemudian evaluasi model dilakukan menggunakan *Confusion Matrix*. Hasil pengujian terbaik diperoleh ketika menggunakan metode *Random Forest* yang dikombinasikan dengan *SMOTE* dan *Chi Square* serta pembagian jumlah data latih dan data uji sebesar 80:20, menghasilkan nilai akurasi sebesar 90,20%, presisi 90,78%, *recall* 90,20% dan *f1-score* 90,17.

Kata Kunci: Aplikasi Grab, Analisis Sentimen, *Text Preprocessing*, *TF-IDF*, *SMOTE*, *Chi Square* dan *Random Forest*.

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN ORISINALITAS.....	iii
KATA PENGANTAR	iv
ABSTRAK	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL	xi
DAFTAR <i>CODE PROGRAM</i>	xii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	4
1.2.1 Permasalahan.....	4
1.2.2 Metode Usulan	4
1.2.3 Pertanyaan Penelitian	4
1.3 Tujuan dan Manfaat	5
1.3.1 Tujuan Penelitian	5
1.3.2 Manfaat Penelitian	5
1.4 Batasan-Batasan	5
1.5 Sistematika Penulisan	6
BAB 2 KAJIAN PUSTAKA.....	7
2.1 Analisis Sentimen	7
2.2 <i>Text Preprocessing</i>	7
2.2.1 <i>Case Folding</i>	7
2.2.2 <i>Cleaning</i>	7
2.2.3 <i>Tokenizing</i>	7
2.2.4 <i>Filtering</i>	8
2.2.5 <i>Stemming</i>	8
2.3 Ekstraksi Fitur <i>TF-IDF</i>	8
2.4 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	9

2.5 Seleksi Fitur <i>Chi Square</i>	10
2.6 Klasifikasi <i>Random Forest</i>	12
2.7 <i>Confusion Matrix</i>	14
2.8 Penelitian Terkait	16
BAB 3 METODE PENELITIAN.....	19
3.1 Pengambilan Data	19
3.2 Seleksi Data.....	20
3.3 Pelabelan Data.....	21
3.4 <i>Text Preprocessing</i>	21
3.4.1 <i>Case Folding</i>	21
3.4.2 <i>Cleaning</i>	22
3.4.3 <i>Tokenizing</i>	23
3.4.4 <i>Filtering</i>	23
3.4.5 <i>Stemming</i>	24
3.5 Ekstraksi Fitur <i>TF-IDF</i>	25
3.6 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	26
3.7 Seleksi Fitur <i>Chi Square</i>	27
3.8 Pembagian Data Latih dan Data Uji.....	32
3.9 Klasifikasi <i>Random Forest</i>	32
3.10 Evaluasi Model.....	35
3.11 Skenario Pengujian.....	36
BAB 4 HASIL DAN PEMBAHASAN.....	37
4.1 Pengambilan Data	37
4.2 Seleksi Data.....	38
4.3 Pelabelan Data.....	39
4.4 <i>Text Preprocessing</i>	40
4.4.1 <i>Case Folding</i>	41
4.4.2 <i>Cleaning</i>	42
4.4.3 <i>Tokenizing</i>	44
4.4.4 <i>Filtering</i>	45
4.4.5 <i>Stemming</i>	46
4.5 Ekstraksi Fitur <i>TF-IDF</i>	48

4.6 <i>Synthetic Minority Oversampling Technique (SMOTE)</i>	49
4.7 Seleksi Fitur <i>Chi Square</i>	50
4.8 Pembagian Data Latih dan Data Uji.....	51
4.9 Klasifikasi <i>Random Forest</i>	53
4.10 Evaluasi Model.....	53
4.11 Hasil Pengujian	56
4.12 Analisa Hasil Pengujian	68
4.13 Implementasi Antarmuka	72
BAB 5 PENUTUP	74
5.1 Kesimpulan	74
5.2 Saran.....	74
DAFTAR PUSTAKA	75
LAMPIRAN.....	78

DAFTAR GAMBAR

Gambar 1. 1 Pengguna <i>mobile operating system</i> di Indonesia (Sumber: <i>gs.statcounter.com</i>)	2
Gambar 3. 1 Arsitektur Sistem.....	19
Gambar 3. 2 Contoh Data Sebelum Proses Seleksi.....	20
Gambar 3. 3 Contoh Data Setelah Proses Seleksi.....	20
Gambar 3. 4 Proses Seleksi Fitur <i>Chi Square</i>	27
Gambar 3. 5 Proses Klasifikasi <i>Random Forest</i>	32
Gambar 3. 6 Proses Pembentukan <i>Subset</i>	33
Gambar 3. 7 Proses Pembentukan <i>Tree</i>	33
Gambar 3. 8 Contoh Perhitungan Nilai Tengah.....	34
Gambar 3. 9 Contoh Perhitungan Nilai Gini dan Gini(<i>split</i>)	34
Gambar 3. 10 <i>Decision Tree</i> dari <i>Subset 1</i>	34
Gambar 3. 11 Proses <i>Majority Voting</i>	35
Gambar 4. 1 Data <i>Output</i> Proses <i>Scraping Data</i>	38
Gambar 4. 2 Data <i>Input</i> Proses Seleksi Data	38
Gambar 4. 3 Data <i>Output</i> Proses Seleksi Data.....	39
Gambar 4. 4 Data <i>Input</i> Proses Pelabelan Data	39
Gambar 4. 5 Data <i>Output</i> Proses Pelabelan Data.....	39
Gambar 4. 6 Hasil Proses <i>Input Data</i>	41
Gambar 4. 7 Data <i>Input</i> Proses <i>Case Folding</i>	42
Gambar 4. 8 Data <i>Output</i> Proses <i>Case Folding</i>	42
Gambar 4. 9 Data <i>Input</i> Proses <i>Cleaning</i>	43
Gambar 4. 10 Data <i>Output</i> Proses <i>Cleaning</i>	43
Gambar 4. 11 Data <i>Input</i> Proses <i>Tokenizing</i>	44
Gambar 4. 12 Data <i>Output</i> Proses <i>Tokenizing</i>	44
Gambar 4. 13 Data <i>Input</i> Proses <i>Filtering</i>	45
Gambar 4. 14 Data <i>Output</i> Proses <i>Filtering</i>	46
Gambar 4. 15 Data <i>Input</i> Proses <i>Stemming</i>	47
Gambar 4. 16 Data <i>Output</i> Proses <i>Stemming</i>	47
Gambar 4. 17 Data <i>Input</i> Proses Ekstraksi Fitur <i>TF-IDF</i>	48

Gambar 4. 18 Data <i>Output</i> Proses Ekstraksi Fitur <i>TF-IDF</i>	49
Gambar 4. 19 Data <i>Input</i> Proses <i>SMOTE</i>	50
Gambar 4. 20 Data <i>Output</i> Proses <i>SMOTE</i>	50
Gambar 4. 21 Data <i>Input</i> Proses Seleksi Fitur <i>Chi Square</i>	51
Gambar 4. 22 Data <i>Output</i> Proses Seleksi Fitur <i>Chi Square</i>	51
Gambar 4. 23 Data <i>Input</i> Proses Pembagian Data	52
Gambar 4. 24 Data <i>Output</i> Proses Pembagian Data (Data Latih).....	52
Gambar 4. 25 Data <i>Output</i> Proses Pembagian Data (Data Uji)	52
Gambar 4. 26 Data <i>Input</i> Proses Klasifikasi <i>Random Forest</i>	53
Gambar 4. 27 Data <i>Output</i> Proses Klasifikasi <i>Random Forest</i>	53
Gambar 4. 28 Data <i>Input</i> <i>Confusion Matrix</i>	55
Gambar 4. 29 Data <i>Output</i> <i>Confusion Matrix</i>	55
Gambar 4. 30 Hasil Pengujian <i>RF</i> 60:40	56
Gambar 4. 31 Hasil Pengujian <i>RF</i> 70:30	57
Gambar 4. 32 Hasil Pengujian <i>RF</i> 80:20	58
Gambar 4. 33 Hasil Pengujian <i>RF+SMOTE</i> 60:40	59
Gambar 4. 34 Hasil Pengujian <i>RF+SMOTE</i> 70:30	60
Gambar 4. 35 Hasil Pengujian <i>RF+SMOTE</i> 80:20	61
Gambar 4. 36 Hasil Pengujian <i>RF+Chi Square</i> 60:40	62
Gambar 4. 37 Hasil Pengujian <i>RF+Chi Square</i> 70:30	63
Gambar 4. 38 Hasil Pengujian <i>RF+Chi Square</i> 80:20	64
Gambar 4. 39 Hasil Pengujian <i>RF+SMOTE+Chi Square</i> 60:40	65
Gambar 4. 40 Hasil Pengujian <i>RF+SMOTE+Chi Square</i> 70:30	66
Gambar 4. 41 Hasil Pengujian <i>RF+SMOTE+Chi Square</i> 80:20	67
Gambar 4. 42 Grafik Hasil Pengujian <i>RF</i>	68
Gambar 4. 43 Grafik Hasil Pengujian <i>RF+SMOTE</i>	69
Gambar 4. 44 Grafik Hasil Pengujian <i>RF+Chi Square</i>	70
Gambar 4. 45 Grafik Hasil Pengujian <i>RF+SMOTE+Chi Square</i>	70
Gambar 4. 46 Grafik Nilai Akurasi Tertinggi.....	71
Gambar 4. 47 Antarmuka Sistem	72
Gambar 4. 48 Contoh Hasil Prediksi Sistem (1)	72
Gambar 4. 49 Contoh Hasil Prediksi Sistem (2)	73

DAFTAR TABEL

Tabel 2. 1 Kontingensi Perhitungan Nilai <i>Chi Square</i>	11
Tabel 2. 2 <i>Confusion Matrix</i>	14
Tabel 2. 3 Rangkuman Penelitian Terkait.....	17
Tabel 3. 1 Contoh Data Relevan dan Tidak Relevan	20
Tabel 3. 2 Contoh Data Setelah Pelabelan	21
Tabel 3. 3 Proses <i>Case Folding</i>	22
Tabel 3. 4 Proses <i>Cleaning</i>	22
Tabel 3. 5 Proses <i>Tokenizing</i>	23
Tabel 3. 6 Proses <i>Filtering</i>	24
Tabel 3. 7 Proses <i>Stemming</i>	24
Tabel 3. 8 Hasil <i>Text Preprocessing</i>	25
Tabel 3. 9 Hasil Perhitungan <i>TF-IDF</i>	26
Tabel 3. 10 Nilai <i>Chi Square</i> Positif	27
Tabel 3. 11 Nilai <i>Chi Square</i> Negatif.....	29
Tabel 3. 12 Nilai <i>Chi Square</i> Tunggal	30
Tabel 3. 13 Nilai <i>Chi Square</i> Tunggal Setelah Diurutkan	31
Tabel 3. 14 Skenario Pengujian	36
Tabel 4. 1 Hasil Proses <i>SMOTE</i>	49
Tabel 4. 2 Hasil Pengujian	68

DAFTAR CODE PROGRAM

<i>Code Program 4. 1 Scraping Data</i>	37
<i>Code Program 4. 2 Import Library</i>	40
<i>Code Program 4. 3 Input Data</i>	40
<i>Code Program 4. 4 Case Folding</i>	41
<i>Code Program 4. 5 Cleaning</i>	42
<i>Code Program 4. 6 Tokenizing</i>	44
<i>Code Program 4. 7 Filtering</i>	45
<i>Code Program 4. 8 Stemming</i>	46
<i>Code Program 4. 9 Ekstraksi Fitur TF-IDF</i>	48
<i>Code Program 4. 10 SMOTE</i>	49
<i>Code Program 4. 11 Chi Square</i>	50
<i>Code Program 4. 12 Pembagian Data Latih dan Data Uji</i>	51
<i>Code Program 4. 13 Klasifikasi Random Forest</i>	53
<i>Code Program 4. 14 Evaluasi Model</i>	54