

KATA PENGANTAR

Puji syukur kepada Tuhan yang Maha Esa atas segala berkat yang telah diberikan-Nya sehingga tugas akhir dengan judul “**PENERAPAN METODE TEXT MINING DENGAN K-MEANS CLUSTERING UNTUK MENGETAHUI SENTIMEN DAN TOPIK POPULER STUDI KASUS: TIGA ONLINE MARKETPLACE DI INDONESIA**” dapat diselesaikan dengan baik.

Tugas Akhir ini disusun untuk memenuhi sebagian persyaratan akademik untuk memperoleh gelar Strata Satu Sarjana Komputer Program Studi Sistem Informasi Universitas Pelita Harapan, Karawaci, Tangerang.

Penulis juga mendapatkan bimbingan, bantuan, dan doa dari berbagai pihak. Karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada pihak yang telah membantu penulis mengerjakan Tugas Akhir ini, yaitu kepada:

- 1) Bapak Dr.Eng., Ir. Pujianto Yugopuspito, M.Sc., selaku Dekan Fakultas Ilmu Komputer.
- 2) Bapak Arnold Aribowo, S.T., M.T., selaku Ketua Program Studi Sistem Informasi yang selalu membantu penulis selama proses perkuliahan maupun pengerjaan Tugas Akhir.
- 3) Bapak Andree Emmanuel Widjaja, Ph.D, selaku Dosen Pembimbing penulis yang selalu memberikan dukungan, masukan, nasihat dan motivasi baik selama proses perkuliahan maupun pengerjaan Tugas Akhir.
- 4) Bapak Arnold Aribowo, S.T., M.T., selaku Pembimbing Akademik penulis yang membantu penulis selama proses perkuliahan maupun pengerjaan Tugas Akhir.
- 5) Seluruh dosen Program Studi Sistem Informasi Universitas Pelita Harapan yang telah mengajar, memberikan motivasi dan membantu penulis selama proses perkuliahan maupun pengerjaan Tugas Akhir.
- 6) Semua staf Tata Usaha dan Karyawan Fakultas Ilmu Komputer Universitas Pelita Harapan yang telah membantu dalam proses administrasi penulis.

- 7) Orang tua dan keluarga yang telah membiayai penulis dalam proses perkuliahan serta memberikan dukungan moral, doa, dan kasih sayang dalam proses perkuliahan sampai dengan menyelesaikan Tugas Akhir ini.
- 8) Kepada teman-teman seperjuangan Amel, Debora, Joshua, Kelvin, Lauren, Wivina yang telah memberikan dukungan dan bantuan kepada penulis.
- 9) Semua teman-teman program studi sistem informasi yang menjadi teman dalam proses perkuliahan.
- 10) Semua pihak yang tidak bisa disebutkan satu per satu.

Akhir kata, penulis menyadari bahwa masih terdapat banyak kekurangan dalam Tugas Akhir ini. Oleh karena itu, kritik dan saran dari pembaca akan sangat bermanfaat bagi penulis. Semoga Tugas Akhir ini dapat bermanfaat bagi semua pihak yang membacanya.

Karawaci, 07 Januari 2021

Penulis

DAFTAR ISI

Pernyataan dan Persetujuan Unggah Tugas Akhir	ii
PERSETUJUAN DOSEN PEMBIMBING TUGAS AKHIR	iii
PERSETUJUAN TIM PENGUJI TUGAS AKHIR	iv
ABSTRAK	v
ABSTRACT	vi
KATA PENGANTAR	vii
DAFTAR ISI	ix
DAFTAR GAMBAR	xii
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah	4
1.3 Pembatasan Masalah	4
1.4 Tujuan dan Manfaat Penelitian	4
1.4.1 Tujuan Penelitian	5
1.4.2 Manfaat Penelitian	5
1.5 Metodologi Penelitian	5
1.5.1 Authentication	5
1.5.2 Data Collecting	6
1.5.3 Text Preprocessing	6
1.5.4 Word Cloud Creating	6
1.5.5 Lexicon	6
1.5.6 Term Weighting	6
1.5.7 Clustering	7
1.6 Sistematika Penulisan	7
BAB II LANDASAN TEORI	9
2.1 <i>Text Mining</i>	9
2.1.1 Cleansing	9
2.1.2 Case Folding	10
2.1.3 Tokenizing	10
2.1.4 Stopwords Removing	10

2.1.5 Stemming	10
2.2 R Studio	11
2.3 Metode <i>Term Frequency – Inverse Document Frequency</i> (TF-IDF)	12
2.4 <i>K-Means Clustering</i>	13
2.5 Twitter.....	14
2.6 <i>Online Marketplace</i>	15
2.7 Penelitian Terdahulu.....	17
BAB III METODOLOGI PENELITIAN	20
3.1 <i>Authentication</i>	20
3.2 <i>Data Collecting</i>	21
3.3 <i>Text Preprocessing</i>	22
3.3.1 Case Folding	22
3.3.2 Stemming	23
3.3.3 Stopwords Removing.....	23
3.4 <i>Word Cloud Creating</i>	24
3.5 <i>Lexicon</i>	24
3.6 <i>Term Weighting</i>	25
3.7 <i>Clustering</i>	25
BAB IV IMPLEMENTASI.....	27
4.1 <i>Authentication</i>	29
4.1.1 Install Package Library Twitter	30
4.1.2 Load Package Library Twitter	30
4.1.3 Integrasi	31
4.2 <i>Data Collecting</i>	31
4.3 <i>Text Preprocessing</i>	36
4.3.1 Case Folding	37
4.3.2 Stemming	40
4.3.3 Stopwords Removing.....	41
4.4 <i>Word Cloud Creating</i>	42
4.5 <i>Lexicon</i>	47
4.6 <i>Term Weighting</i>	50
4.7 <i>Clustering</i>	57
4.8 Analisa	65
BAB V KESIMPULAN DAN SARAN.....	71

5.1 Kesimpulan.....	71
5.2 Batasan Penelitian dan Saran.....	71
DAFTAR PUSTAKA.....	73



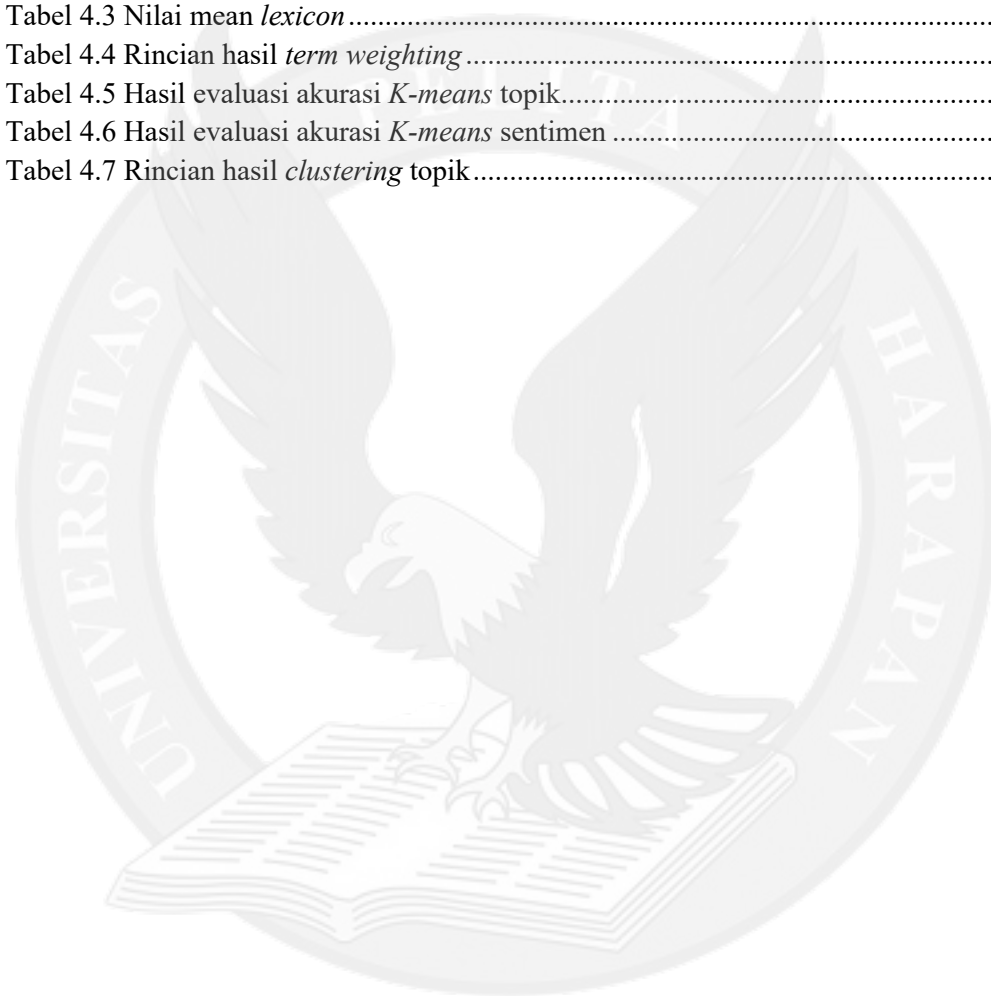
DAFTAR GAMBAR

Gambar 1.1 Sepuluh negara dengan pertumbuhan <i>e-commerce</i> tercepat.....	2
Gambar 2.1 Contoh alur proses <i>text preprocessing</i>	11
Gambar 2.2 Grafik pertumbuhan <i>e-commerce</i> di Indonesia.....	16
Gambar 3.1 Skema proses <i>authentication</i>	20
Gambar 3.2 Skema proses <i>data collecting</i>	21
Gambar 3.3 Grafik pengunjung terbesar <i>e-commerce</i> di Indonesia	22
Gambar 3.4 Metodologi yang digunakan	26
Gambar 4.1 Jenis-jenis <i>software</i> R studio	27
Gambar 4.2 Bagian <i>software</i> R studio.....	28
Gambar 4.3 Kode <i>install package library twitter</i>	30
Gambar 4.4 Kode <i>load package library twitter</i>	30
Gambar 4.5 Kode integrasi R studio dan twitter	31
Gambar 4.6 Kode pengambilan <i>tweets</i> tahap 1	32
Gambar 4.7 Kode pengambilan <i>tweets</i> tahap 2	32
Gambar 4.8 Kode pengambilan <i>tweets</i> tahap 3	32
Gambar 4.9 Kode fungsi <i>dataframe</i>	33
Gambar 4.10 Kode penyimpanan CSV	34
Gambar 4.11 Kode <i>read file CSV</i>	35
Gambar 4.12 Kode <i>install package dplyr</i>	35
Gambar 4.13 Kode penggabungan file CSV	35
Gambar 4.14 Kode penyimpanan file CSV.....	36
Gambar 4.15 Kode <i>install package text preprocessing</i>	37
Gambar 4.16 Kode <i>read file text preprocessing</i>	38
Gambar 4.17 Kode <i>case folding</i> tahap 1.....	38
Gambar 4.18 Kode <i>case folding</i> tahap 2.....	39
Gambar 4.19 Kode <i>case folding</i> tahap 3.....	39
Gambar 4.20 Kode <i>case folding</i> tahap 4.....	40
Gambar 4.21 Kode <i>case folding</i> tahap 5.....	40
Gambar 4.22 Kode <i>stemming</i>	41
Gambar 4.23 Kode <i>stopwords removing</i> tahap 1	41
Gambar 4.24 Kode <i>stopwords removing</i> tahap 2	42
Gambar 4.25 Kode <i>word cloud creating</i> tahap 1.....	42
Gambar 4.26 Kode <i>word cloud creating</i> tahap 2.....	43
Gambar 4.27 Kode <i>word cloud creating</i> tahap 3.....	43
Gambar 4.28 Kode <i>word cloud creating</i> tahap 4.....	44
Gambar 4.29 Hasil <i>word cloud</i> gabungan	45
Gambar 4.30 Hasil <i>word cloud</i> Tokopedia.....	45
Gambar 4.31 Hasil <i>word cloud</i> Shopee	46
Gambar 4.32 Hasil <i>word cloud</i> Bukalapak.....	46
Gambar 4.33 Kode <i>lexicon</i> tahap 1	47

Gambar 4.34 Kode <i>lexicon</i> tahap 2	48
Gambar 4.35 Kode <i>lexicon</i> tahap 3	48
Gambar 4.36 Kode <i>lexicon</i> tahap 4	49
Gambar 4.37 Kode <i>lexicon</i> tahap 5	49
Gambar 4.38 Kode <i>lexicon</i> tahap 6	49
Gambar 4.39 Kode <i>lexicon</i> tahap 7	50
Gambar 4.40 Kode <i>term weighting</i> tahap 1	50
Gambar 4.41 Kode <i>term weighting</i> tahap 2	51
Gambar 4.42 Kode <i>term weighting</i> tahap 3	51
Gambar 4.43 Kode <i>term weighting</i> tahap 4	52
Gambar 4.44 Hasil <i>term frequency</i> gabungan	52
Gambar 4.45 Hasil <i>term frequency</i> Tokopedia.....	53
Gambar 4.46 Hasil <i>term frequency</i> Shopee	53
Gambar 4.47 Hasil <i>term frequency</i> Bukalapak.....	54
Gambar 4.48 Kode <i>term weighting</i> tahap 5	54
Gambar 4.49 Kode <i>term weighting</i> tahap 6	55
Gambar 4.50 Hasil TF-IDF gabungan.....	55
Gambar 4.51 Hasil TF-IDF Tokopedia	56
Gambar 4.52 Hasil TF-IDF Shopee.....	56
Gambar 4.53 Hasil TF-IDF Bukalapak	57
Gambar 4.54 Kode <i>clustering</i> tahap 1	58
Gambar 4.55 Kode <i>clustering</i> tahap 2	58
Gambar 4.56 Hasil <i>K-means</i> topik gabungan.....	59
Gambar 4.57 Hasil <i>K-means</i> topik Tokopedia	60
Gambar 4.58 Hasil <i>K-means</i> topik Shopee.....	60
Gambar 4.59 Hasil <i>K-means</i> topik Bukalapak	61
Gambar 4.60 Evaluasi <i>K-means</i> topik	61
Gambar 4.61 Kode <i>clustering</i> tahap 3	62
Gambar 4.62 Kode <i>clustering</i> tahap 4	62
Gambar 4.63 Kode <i>clustering</i> tahap 5	62
Gambar 4.64 Hasil <i>K-Means</i> sentimen gabungan	63
Gambar 4.65 Hasil <i>K-Means</i> sentimen Tokopedia.....	63
Gambar 4.66 Hasil <i>K-Means</i> sentimen Shopee	64
Gambar 4.67 Hasil <i>K-Means</i> sentimen Bukalapak.....	64
Gambar 4.68 Evaluasi <i>K-means</i> sentimen	65
Gambar 4.68 Daftar negara dengan ciutan <i>kpop</i> terbanyak di twitter	69

DAFTAR TABEL

Tabel 2.1 Penelitian terdahulu.....	17
Tabel 3.1 Pembagian <i>data collecting</i>	21
Tabel 3.2 Contoh <i>stopwords</i>	24
Tabel 3.3 Penilaian sentimen.....	24
Tabel 4.1 Pembagian <i>data collecting</i>	31
Tabel 4.2 Rangkuman hasil <i>word cloud</i>	47
Tabel 4.3 Nilai mean <i>lexicon</i>	50
Tabel 4.4 Rincian hasil <i>term weighting</i>	57
Tabel 4.5 Hasil evaluasi akurasi <i>K-means</i> topik.....	61
Tabel 4.6 Hasil evaluasi akurasi <i>K-means</i> sentimen	65
Tabel 4.7 Rincian hasil <i>clustering</i> topik.....	67



DAFTAR LAMPIRAN

LAMPIRAN A DATA <i>TWEETS</i> TOKOPEDIA.....	A-1
LAMPIRAN B DATA <i>TWEETS</i> SHOPEE.....	B-1
LAMPIRAN C DATA <i>TWEETS</i> BUKALAPAK	C-1
LAMPIRAN D DAFTAR <i>STOPWORDS</i>	D-1

