

DAFTAR ISI

	Hlm.
HALAMAN JUDUL	
LEMBAR PERSETUJUAN	
LEMBAR PENGESAHAN	
LEMBAR PERNYATAAN KARYA SENDIRI	
LEMBAR PERSEMBAHAN	
ABSTRAK	i
ABSTRACT	ii
KATA PENGANTAR	iii
DAFTAR ISI	v
DAFTAR GAMBAR	viii
DAFTAR TABEL	x
DAFTAR LAMPIRAN	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	5
1.6 Kerangka Pemikiran	5
1.7 Sistematika Penulisan	6
BAB II KAJIAN LITERATUR	8
2.1 Tinjauan Pustaka	8
2.2 Landasan Teori	14
2.2.1 Web Scraping	14
2.2.2 Text Preprocessing	14
2.2.3 N-Gram	15
2.2.4 Term Frequency	15
2.2.5 Cosine Similarity	16
2.2.6 Frequence Ask Question	16
2.2.7 Question Generation	17
2.2.8 Instagram	17
2.2.9 Instaloader	17

2.2.10	Bahasa Pemrograman <i>Python</i>	17
2.2.11	<i>Flowchart</i>	18
2.2.12	<i>CRISP-DM</i>	19
BAB III METODOLOGI PENELITIAN		21
3.1	Pemahaman Bisnis (<i>Bussiness Understanding</i>).....	21
3.2	Pemahaman Data (<i>Data Understanding</i>).....	21
3.3	Persiapan Data (<i>Data Preparation</i>).....	23
3.3.1.	<i>Cleansing</i>	24
3.3.2.	<i>Case Folding</i>	24
3.3.3.	<i>Stopwords Removal</i>	25
3.3.4.	<i>Tokenizing</i>	25
3.3.5.	<i>Stemming</i>	26
3.4	Pemodelan (<i>Modeling</i>)	27
3.4.1.	<i>N-Gram</i>	27
3.4.2.	Pembangkit Pertanyaan	28
3.4.3.	<i>Replace</i>	29
3.4.4.	<i>Term Frequency</i>	30
3.5	Evaluasi (<i>Evaluation</i>)	30
3.5.1	<i>Cosine Similarity</i>	30
3.6	<i>Deployment</i>	31
3.6.1	Analisis Kebutuhan Perangkat Keras (<i>Hardware</i>).....	32
3.6.2	Analisis Kebutuhan Perangkat Lunak (<i>Software</i>)	32
BAB IV HASIL DAN PEMBAHASAN		33
4.1	Pembangkitan Pertanyaan Berdasarkan Data.....	33
4.1.1	Hasil Pemahaman Bisnis	33
4.1.2	Hasil Pemahaman Data	34
4.1.3	Hasil Persiapan Data	39
4.1.4	Hasil Pemodelan.....	47
4.2	Kinerja <i>Cosine Similarity</i>	54
4.2.1	Hasil Evaluasi	54
4.3	Tampilan Pertanyaan Pada Web.....	61
4.3.1	<i>Deployment</i>	61
BAB V KESIMPULAN DAN SARAN		64
5.1	Kesimpulan.....	64

5.2	Saran	64
	DAFTAR PUSTAKA	xii
	LAMPIRAN	xvi



DAFTAR GAMBAR

	Hlm.
Gambar 2. 1 Alur Tahapan <i>CRISP-DM</i>	19
Gambar 3. 1 Alur Flowchart Pada Tahap Pemahaman Data	22
Gambar 3. 2 Alur Flowchart Pada Tahap Persiapan Data	23
Gambar 3. 3 Alur Flowchart Pada Tahap Pemodelan	27
Gambar 4. 1 Alur <i>Flowchart</i> pada tahap pemahaman bisnis	34
Gambar 4. 2 Proses <i>Crawling Dataset</i>	35
Gambar 4. 3 Sampel Hasil <i>Crawling Dataset</i>	36
Gambar 4. 4 <i>Install Package</i> yang Dibutuhkan	36
Gambar 4. 5 Impor <i>Library</i> yang Dibutuhkan	37
Gambar 4. 6 Proses Penggabungan Data	37
Gambar 4. 7 Proses Pengambilan Data Teks dan Answer	38
Gambar 4. 8 <i>Preview Data</i>	38
Gambar 4. 9 Fungsi Pada Proses <i>Cleansing</i>	39
Gambar 4. 10 Fungsi Pada Proses <i>Cleansing</i>	40
Gambar 4. 11 Fungsi Pada Proses <i>Cleansing</i>	40
Gambar 4. 12 Proses Implementasi Fungsi <i>Cleansing</i>	41
Gambar 4. 13 <i>Output</i> Tahapan <i>Cleansing</i>	42
Gambar 4. 14 Impor <i>Library Stopwords</i>	42
Gambar 4. 15 Fungsi Pada Proses <i>Stopwords Removal</i>	43
Gambar 4. 16 Proses Implementasi <i>Stopwords Removal Term Frequency</i>	43
Gambar 4. 17 Proses Implementasi <i>Stopwords Removal N-Gram</i>	43
Gambar 4. 18 <i>Output</i> Tahapan <i>Stopwords Removal Term Frequency</i>	44
Gambar 4. 19 <i>Output</i> Tahapan <i>Stopwords Removal N-Gram</i>	44
Gambar 4. 20 Fungsi Pada Proses <i>Stemming</i>	45
Gambar 4. 21 Proses <i>Stemming N-Gram</i>	45
Gambar 4. 22 Proses <i>Stemming Term Frequency</i>	45
Gambar 4. 23 <i>Output</i> Tahapan <i>Stemming N-Gram</i>	46
Gambar 4. 24 <i>Output</i> Tahapan <i>Stemming Term Frequency</i>	46
Gambar 4. 25 Proses <i>Bi-Gram</i>	47
Gambar 4. 26 Proses Filterisasi Hasil <i>Bi-Gram</i>	47
Gambar 4. 27 <i>Output</i> Tahapan <i>Bi-Gram</i>	48
Gambar 4. 28 Proses Pencarian dan Penggabungan Data	48
Gambar 4. 29 Sampel Hasil Pencarian Data Berdasarkan <i>Bi-Gram</i>	49
Gambar 4. 30 <i>Output Uni-Gram</i>	49
Gambar 4. 31 <i>Output Tri-Gram</i>	50
Gambar 4. 32 Fungsi Pada Proses <i>Replace</i>	50
Gambar 4. 33 <i>Output</i> Tahapan <i>Replace</i>	51
Gambar 4. 34 Proses <i>Term Frequency</i>	52
Gambar 4. 35 <i>Output</i> Tahapan <i>Term Frequency</i>	52
Gambar 4. 36 Proses Pembangkit Pertanyaan	53

Gambar 4. 37 <i>Output</i> Tahapan Pembangkit Pertanyaan	54
Gambar 4. 38 Proses <i>Cosine Similarity</i>	55
Gambar 4. 39 Proses <i>Deployment</i>	62
Gambar 4. 40 Hasil <i>Deployment</i> Pada Tampilan Admin	62
Gambar 4. 41 Hasil <i>Deployment</i> Pada Tampilan Publik.....	63



DAFTAR TABEL

	Hlm.
Tabel 2. 1 <i>State of The Art</i>	10
Tabel 2. 2 Simbol <i>Flowchart</i>	18
Tabel 3. 1 Sampel Dataset.....	21
Tabel 3. 2 Sampel Hasil <i>Cleansing</i>	24
Tabel 3. 3 Sampel Hasil <i>Case Folding</i>	24
Tabel 3. 4 Tambahan Daftar Kata <i>Stopwords Removal</i>	25
Tabel 3. 5 Sampel Hasil <i>Stopwords Removal</i>	25
Tabel 3. 6 Sampel Hasil <i>Tokenizing</i>	26
Tabel 3. 7 Sampel Hasil <i>Stemming</i>	26
Tabel 3. 8 Sampel Hasil <i>Text Preprocessing</i>	27
Tabel 3. 9 Sampel Hasil <i>Bigram</i>	28
Tabel 3. 10 Sampel Komentar yang Mengandung <i>Bigram</i>	28
Tabel 3. 11 Sampel Hasil Pembangkit Pertanyaan.....	28
Tabel 3. 12 Sampel Hasil <i>Replace</i>	29
Tabel 3. 13 Sampel Hasil <i>Term Frequency</i>	30
Tabel 3. 14 Hasil Pembangkit Pertanyaan dan Term Frequency.....	30
Tabel 3. 15 Hasil Perhitungan Cosine Similarity	31
Tabel 3. 16 Kebutuhan Perangkat Keras	32
Tabel 3. 17 Kebutuhan Perangkat Lunak	32
Tabel 4. 1 Hasil <i>Cosine Similarity</i> dengan Kata “daftar wisuda”	55
Tabel 4. 2 Hasil <i>Cosine Similarity</i> dengan Kata “bayar ukt”	57
Tabel 4. 3 Hasil <i>Cosine Similarity</i> dengan Kata “jalur mandiri”	58
Tabel 4. 4 Hasil <i>Cosine Similarity</i> dengan Kata “jalur spanptkin”	59
Tabel 4. 5 Hasil <i>Cosine Similarity</i> dengan Kata “ujian mandiri”	59
Tabel 4. 6 Hasil <i>Cosine Similarity</i> dengan Kata “jalur snbp”	60
Tabel 4. 7 Hasil <i>Cosine Similarity</i> dengan Kata “jalur umptkin”	60

DAFTAR LAMPIRAN

	Hlm.
Lampiran A <i>Source code Web Scraping</i>	xvi
Lampiran B <i>Source code Cleansing</i>	xvi
Lampiran C <i>Source code Stopwords Removal</i>	xviii
Lampiran D <i>Source code Stemming</i>	xix
Lampiran E <i>Source code N-Gram</i>	xix
Lampiran F <i>Source code Replace</i>	xx
Lampiran G <i>Source code Question Generation</i>	xx
Lampiran H <i>Source code Term Frequency</i>	xxi
Lampiran I <i>Source code Cosine Similarity</i>	xxii
Lampiran J <i>Source code Deployment</i>	xxii
Lampiran K <i>Source code index.html</i>	xxiii
Lampiran L <i>Source code refactor.py</i>	xxv
Lampiran M Gambar Hasil Cosine Similarity	xxx