PERINGKASAN TEKS BAHASA INDONESIA OTOMATIS MENGGUNAKAN METODE FEEDFORWARD NEURAL NETWORK

(Skripsi)

Oleh :

ANNISA NUR FADHILAH

JURUSAN ILMU KOMPUTER

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS LAMPUNG

2018

ABSTRACT

AUTOMATIC TEXT SUMMARIZATION IN BAHASA INDONESIA USING FEEDFORWARD NEURAL NETWORK

By

ANNISA NUR FADHILAH

This research is intended to display the weighting of text features to summarize documents in Indonesian using feedforward neural network (FFNN). This study involves eleven text features taken from the research of Aristoteles et al (2012), the position phrase (f_1), positive keywords in sentence (f_2), negative keywords in sentences (f_3), sentence centrality (f_4), sentence resemblance to the title (f_5), sentence inclusion of name entity (f_6), sentence inclusion of numerical data (f_7), relative length of sentences (f_8), thick lane nodes (f_9), summing similarities for each node (f_{10}), and latent semantic features (f_{11}). All feature scoring functions are used to train the FFNN model to obtain a combination of the appropriate feature weights. Evaluation of text summarization using F-measure. The results showed average F-measure 90.9% for compression rate 30%, 76.1% for compression rate 20%, and 49.1% for compression rate 10%. Just using f_2 can provide the same performance by using the eleventh feature. The computer specifications used affect the computing time required for a single runnning of the program.

Keywords : Summarization Text, Feedforward Neural Network, Latent Semantic Feature.

ABSTRAK

PERINGKASAN TEKS BAHASA INDONESIA OTOMATIS MENGGUNAKAN METODE FEEDFORWARD NEURAL NETWORK

Oleh

ANNISA NUR FADHILAH

Penelitian ini menampilkan bobot fitur teks untuk meringkas dokumen dalam bahasa Indonesia menggunakan *feedforward neural network* (FFNN). Penelitian ini melibatkan sebelas fitur teks yang diambil dari penelitian Aristoteles et al (2012), yaitu posisi kalimat (f_1), *positive keyword* (f_2), *negative keyword* (f_3), kemiripan antar kalimat (f_4), kalimat yang menyerupai judul kalimat (f_5), kalimat yang mengandung nama entity (f_6), kalimat yang mengandung data numerik (f_7), panjang kalimat (f_8), koneksi antar kalimat (f_9), dan penjumlahan bobot koneksi antar-kalimat (f_{10}) , kalimat semantik (f_{11}). Skor setiap kalimat yang diperoleh digunakan untuk melatih model FFNN sehingga didapatkan kombinasi bobot fitur yang optimal. Evaluasi hasil peringkasan teks yang dihasilkan sistem menggunakan metode *F-measure*. Hasil penelitian menunjukkan rata-rata *F-measure* 90,9% untuk tingkat kompresi 30%, 76,1% untuk tingkat kompresi 20%, dan 49,1% untuk tingkat kompresi 10%. Berdasarkan analisa model bobot, fitur teks f_2 merupakan fitur yang paling berpengaruh dan dapat mewakili kesepuluh fitur lainya. Spesifikasi komputer yang digunakan mempengaruhi waktu komputasi yang diperlukan untuk mengeksekusi program peringkasan teks.

Kata kunci : Peringkasan Teks, *Feedforward Neural Network*, Fitur Kalimat Semantik.

PERINGKASAN TEKS BAHASA INDONESIA OTOMATIS MENGGUNAKAN METODE FEEDFORWARD NEURAL NETWORK

Oleh

ANNISA NUR FADHILAH

Skripsi

Sebagai Salah Satu Syarat untuk Memperoleh Gelar
SARJANA KOMPUTER

pada

Jurusan Ilmu Komputer
Fakultas Matematika dan Ilmu Pengetahuan Alam

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMPUNG
BANDAR LAMPUNG
2018

Judul Skripsi	:	**PERINGKASAN TEKS BAHASA INDONESIA OTOMATIS MENGGUNAKAN METODE FEEDFORWARD NEURAL NETWORK**
Nama Mahasiswa	:	Annisa Nur Fadhilah
No. Pokok Mahasiswa	:	1317051010
Jurusan	:	Ilmu Komputer
Fakultas	:	Matematika dan Ilmu Pengetahuan Alam

MENYETUJUI

1. Komisi Pembimbing

Aristoteles, S.Si., M.Si.	**Rico Andrian, S.Si., M.Kom.**
NIP 19810521 200604 1 002	NIP 19750627 200501 1 001

2. Mengetahui
Ketua Jurusan Ilmu Komputer
FMIPA Universitas Lampung

Dr. Ir. Kurnia Muludi, M.S.Sc.
NIP 19640616 198902 1 001

MENGESAHKAN

1. Tim Penguji

   Ketua : **Aristoteles, S.Si., M.Si.** .....

   Sekretaris : **Rico Andrian, S.Si., M.Kom.** .....

   Penguji
   Bukan Pembimbing : **Dr. Eng. Admi Syarif** .....

2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

Prof. Warsito, S.Si. D.E.A., Ph.D.
NIP. 19710212 199512 1 001

Tanggal Lulus Ujian Skripsi : **15 Januari 2018**

PERNYATAAN

Saya yang bertanda tangan di bawah ini, menyatakan bahwa skripsi saya yang berjudul "Peringkasan Teks Bahasa Indonesia Otomatis Menggunakan Metode *Feedforward Neural Network*" merupakan karya saya sendiri dan bukan karya orang lain. Semua tulisan yang tertuang di skripsi ini telah mengikuti kaidah penulisan karya ilmiah Universitas Lampung. Apabila di kemudian hari terbukti skripsi saya merupakan hasil penjiplakan atau dibuat orang lain, maka saya bersedia menerima sanksi berupa pencabutan gelar yang telah saya terima.

Bandar Lampung, Januari 2018

ANNISA NUR FADHILAH

NPM 1317051010

RIWAYAT HIDUP

Penulis dilahirkan pada tanggal 09 Januari 1996 di Bandar Lampung, merupakan anak kedua dari tiga bersaudara. Penulis menyelesaikan pendidikan formal pertama kali di Taman Kanak-kanak (TK) Dharma Wanita Persatuan, Sukarame pada tahun 2001, menyelesaikan Sekolah Dasar (SD) di SD Negeri 2 Harapan Jaya, Sukarame, Bandar Lampung pada tahun 2007, menyelesaikan Sekolah Menengah Pertama (SMP) di SMP Negeri 23 Bandar Lampung pada tahun 2010, kemudian melanjutkan jenjang Sekolah Menengah Atas (SMA) di MAN 1 Bandar Lampung dan lulus pada tahun 2013.

Pada tahun 2013, Penulis terdaftar sebagai mahasiswi Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung melalui jalur SBMPTN. Pada bulan Januari – Maret 2016, Penulis melakukan Kerja Praktik (KP) di Dinas Kesehatan Kota Bandar Lampung selama 40 hari. Kemudian pada bulan Juli 2016, Penulis melaksanakan Kuliah Kerja Nyata (KKN) selama 40 hari di Desa Kediri, Kecamatan Gadingrejo, Kabupaten Pringsewu.

Selama menjadi mahasiswi, beberapa kegiatan yang pernah dilakukan Penulis antara lain :

1. Aktif dalam Organisasi Himpunan Mahasiswa Jurusan Ilmu Komputer (Himakom) Universitas Lampung dengan menjabat sebagai Anggota Bidang Kesekretariatan pada tahun 2015-2016.
2. Aktif dalam Organisasi Rohani Islam (ROIS) FMIPA Universitas Lampung dengan menjabat sebagai Anggota Bidang Informasi dan Komunikasi (Infokom) pada tahun 2014-2015 dan sebagai Sekretaris Bidang Infokom pada tahun 2015-2016.
3. Aktif dalam Organisasi Badan Eksekutif Mahasiswa (BEM) FMIPA Universitas Lampung dengan menjabat sebagai Sekretaris Departemen Komunikasi dan Informasi (Kominfo) pada tahun 2016.

PERSEMBAHAN

Puji dan syukur saya panjatkan kepada Allah SWT atas segala berkah-Nya yang telah diberikan kepada saya sehingga skripsi ini dapat terselesaikan. Dengan segala kerendahan hati, kupersembahkan karya kecil ini kepada :

Yang tersayang, Bunda. Malaikat tak bersayap, anugrah terbesar yang dititipkan Allah dalam hidup.

Kakak dan adik, Anandyta Nur Khoirunnisa dan Muhammad Arief Islamy.

Keluarga Ilmu Komputer 2013

Serta Almamater tercinta, Universitas Lampung.

MOTTO

“Sesungguhnya sesudah kesulitan itu ada kemudahan.”

(Q.S. Al-Insyiroh : 6)

“There is no secret to be success. It just about doing your best everyday.”

(Lin Dan)

“Nobody guarantees you will succeed if you don’t give up. But if you do give up, there will be nothing left.”

(Aomine Daiki)

SANWACANA

Assalamu'alaikum Wr. Wb.

Alhamdulillah, segala puji bagi Allah SWT yang telah melimpahkan segala nikmat, sehingga Penulis dapat menyelesaikan penulisan skripsi yang berjudul "Peringkasan Teks Bahasa Indonesia Otomatis Menggunakan Metode *Feedforward Neural Network*" dengan baik. Terima kasih Penulis ucapkan kepada semua pihak yang telah membantu dan berperan besar dalam menyusun skripsi ini, antara lain :

1. Seluruh keluarga yang telah memberikan dukungan setiap waktu dan menjadi motivasi terbesar bagi Penulis.
2. Bapak Aristoteles, S.Si., M.Si., sebagai pembimbing utama yang telah dengan sabar membimbing, memotivasi, serta memberikan ide, kritik, dan saran selama masa perkuliahan, KP, dan penyusunan skripsi.
3. Bapak Rico Andrian, S.Si., M.Kom., sebagai pembimbing kedua yang telah dengan sabar membimbing, memberikan ide, kritik, dan saran dalam penyusunan skripsi ini.
4. Bapak Dr. Eng. Admi Syarif sebagai pembahas yang telah memberikan banyak kritik dan saran yang bermanfaat untuk perbaikan dalam penyusunan skripsi ini.
5. Bapak Prof. Warsito, S.Si., D.E.A., Ph.D., sebagai Dekan FMIPA Universitas Lampung.
6. Bapak Dr. Ir. Kurnia Muludi, M.S.Sc., sebagai Ketua Jurusan Ilmu Komputer FMIPA Universitas Lampung.
7. Bapak Didik Kurniawan, S.Si., M.T., sebagai Sekretaris Jurusan Ilmu Komputer FMIPA Universitas Lampung yang telah banyak membantu penulis selama perkuliahan.
8. Bapak dan Ibu Dosen Jurusan Ilmu Komputer yang telah memberikan ilmu selama Penulis menjadi mahasiswa.

9. Sahabat tercinta, Afifah Qowwamina Qisty, Fitria Ramadhani, Qory Aprilarita, Yeni Nuh Richa Sari, Dini Khansa Al-Nakhaiyah, Eria Ayu Ningtias, Agung Prasetyo, Faiq Sulthon Dani, Aulia Rahmawati, Annisa Hasta Pratiwi, Laras Aulia Putri, Briliana Suryani Kusumakinasih, Desy Kartika Sari, terima kasih atas semua doa dan dukungan yang diberikan dalam setiap keadaan.
10. Teman-teman seperjuangan, Irfani Maharani, Putri Dwi Pangestu, Rita Dwilestari, Bunga Ridho Eldita, Widi Novian N, Rifal Kasa Dinar.
11. Ibu Triningrum, SKM., terima kasih atas doa yang diberikan kepada Penulis.
12. Keluarga KKN Desa Kediri, Mbah Saring, Mbah Turni, Pak Aan, Bude Nani, Ifa, Citra, Azmi Prilly Naisa, Citra Alam P, Mbak Dea Raissa, Ezra Rheinsky Tiarsa, Kak Sidik Aryono, Romario Sihombing.
13. Keluarga besar ROIS FMIPA Unila, terima kasih atas pengalaman dan ukhuwah yang terjalin sampai saat ini.
14. Kominfo BEM FMIPA 2016 dan Tim 3D KWI 2017, Aatin, Adji, Ami, Ani, Asih, Elma, Fahrul, Farida, Inas, Mahyal, Nurul, Raka, Rizky, Vina, Ratri, Tamara, Yohanes, Yudha, Yuni, Ani Nurhayati, Rofi, Azizah, Bagus, Astri, Rajes, Dwi, Intan, Resvy, Rizka, Yola, Yomi, dan Putri, terima kasih atas pengalaman dan kebersamaan singkat yang sangat berkesan.

Semoga Allah memberikan balasan terbaik atas semua kebaikan yang telah diberikan kepada Penulis. Semoga skripsi ini dapat bermanfaat bagi orang lain dan perkembangan ilmu pengetahuan.

Bandar Lampung, Januari 2018
Penulis,

Annisa Nur Fadhilah
1317051010

DAFTAR ISI

Halaman

I. PENDAHULUAN 1

1.1 Latar Belakang 1

1.2 Rumusan Masalah 3

1.3 Batasan Masalah 4

1.4 Tujuan 4

1.5 Manfaat 4

II. TINJAUAN PUSTAKA 5

2.1 Peringkasan Teks 5

2.2 Ekstraksi Fitur Teks 5

2.2.1 Posisi Kalimat (f_1) 6

2.2.2 *Positive Keyword* (f_2) 7

2.2.3 *Negative Keyword* ($f3$) 9

2.2.4 Kemiripan Antar-Kalimat ($f4$) 10

2.2.5 Kalimat yang Menyerupai Judul Dokumen (f_5) 11

2.2.6 Kalimat yang Mengandung Nama Entiti (f_6) 12

2.2.7 Kalimat yang Mengandung Data Numerik (f_7) 13

2.2.8 Panjang Kalimat (f_8) 14

2.2.9 Koneksi Antar-Kalimat (f_9) 16

2.2.10 Penjumlahan Bobot Koneksi Antar-Kalimat (f_{10}) 17

2.2.11 Kalimat Semantik (f_{11}) 17

2.3 Sejarah Jaringan Syaraf Tiruan 21

2.3.1 Dasar Jaringan Syaraf Tiruan 23

2.3.2 Model Dasar Jaringan Syaraf Tiruan 23

2.3.3 Feedforward Neural Network (FFNN) 24

2.3.4 Arsitektur Jaringan 24

2.3.5 Neuron 26

2.3.6 Perceptron 27
2.3.7 Paradigma Pembelajaran 27
2.3.8 Fungsi Aktivasi 28
2.3.9 Bias 31
III. METODE PENELITIAN 32
3.1 Prosedur Penelitian 32
3.1.1 Tahap Pengumpulan Dokumen 33
3.1.2 Tahap Pelatihan 33
3.1.3 Tahap Pengujian 34
3.2 Lingkungan Pengembangan 35
3.3 Lingkungan Pengembangan 35
IV. HASIL DAN PEMBAHASAN 36
4.1 Format Dokumen 36
4.2 Data Korpus 36
4.3 Implementasi Sistem 37
4.4 Pemodelan Algoritma Feedforward Neural Network (FFNN) 38
4.5 Hasil Percobaan Model Algoritma Feedforward Neural Network 41
4.6 Waktu Komputasi 43
4.7 Analisa Model Bobot 44
4.8 Perbandingan FFNN dengan Algoritma Genetika 45
4.9 Perbandingan FFNN dengan Regresi Logistik Biner 46
V. KESIMPULAN DAN SARAN 48
5.1 Kesimpulan 48
5.2 Saran 49
DAFTAR PUSTAKA 50

DAFTAR TABEL

Tabel Halaman

1. 1 Hasil percobaan yang dilakukan (Fattah & Ren, 2008) 2
4. 1 Data akurasi pelatihan 41
4. 2 Data akurasi pengujian 42
4. 3 Analisa model bobot 45
4. 4 Perbandingan akurasi FFNN dan GA berdasarkan pemampatan 46
4. 5 Perbandingan akurasi FFNN dan Regresi Logistik Biner (CR 30%) 47

DAFTAR GAMBAR

Gambar Halaman

2. 1 Ilustrasi perhitungan skor kalimat semantik 19
2. 2 Jaringan Syaraf Tiruan *FeedForward* 25
2. 3 *Single Layer Neural Network* 26
2. 4 *Multilayer Network* 26
2. 5 Bentuk Peceptron 27
2. 6 Fungsi Aktivasi 29
2. 7 Jaringan Syaraf Tiruan dengan Bias 31
3. 1 Diagram Alir Tahap Pelatihan dan Pengujian Peringkasan Teks …..…….. 32
4. 1 Tahapan mengolah dokumen korpus ……...……. 37
4. 2 Pemodelan FFNN 39
4. 3 Grafik konsistensi akurasi pengujian dengan pemampatan 30% 42
4. 4 Grafik konsistensi akurasi pengujian dengan pemampatan 20% 42
4. 5 Grafik konsistensi akurasi pengujian dengan pemampatan 10% 42
4. 6 Perbandingan akurasi dengan pemampatan 30%, 20%, dan 10% 43
4. 7 Perbandingan waktu komputasi berdasarkan spesifikasi komputer 43
4. 8 Perbandingan waktu komputasi berdasarkan variasi pemampatan 44

I. PENDAHULUAN

1.1 Latar Belakang

Peringkasan teks merupakan salah satu pengembangan sistem cerdas yang menjadi fokus banyak penelitian dalam beberapa tahun terakhir. Menurut (Jolly & Chopra, 2016) alasan peringkasan teks menjadi sangat penting untuk dikembangkan adalah untuk menghemat waktu pembacaan dokumen teks berskala besar, memudahkan pemilihan dokumen dalam pencarian literatur, meningkatkan efisiensi dalam pengindeksan dokumen, kemungkinan bias mendekati nol, memudahkan dalam menjawab pertanyaan terkait intisari dokumen, dan dapat memudahkan dalam melakukan evaluasi teks yang telah dipublikasi. Menurut (Aristoteles et al., 2012), peringkasan secara manual terhadap dokumen berskala besar akan membutuhkan banyak waktu dan biaya, sehingga peringkasan teks secara otomatis diperlukan untuk menghindari masalah tersebut. Peringkasan teks merupakan sebuah proses untuk menemukan intisari dari dokumen berskala besar dengan mengambil kalimat-kalimat yang dianggap penting. Ada banyak metode yang dapat diterapkan dalam proses ini. Beberapa penelitian yang telah dilakukan sebelumnya dengan berbagai metode menunjukkan hasil akurasi yang beragam.

Penelitian peringkasan teks yang dilakukan oleh (Fattah & Ren, 2008) melibatkan sepuluh fitur teks dengan membandingkan metode Algoritma Genetika dan metode Regresi Matematika. Rincian hasil percobaan dapat dilihat pada Tabel 1.1.

Tabel 1. 1 Hasil percobaan yang dilakukan (Fattah & Ren, 2008)

Pemampatan / Metode	Algoritma Genetika	Regresi Matematika
10%	43.76%	43.12%
20%	44.35%	43.53%
30%	44.94%	43.82%

Penelitian yang dilakukan oleh (Aristoteles et al., 2012) adalah peringkasan teks menggunakan metode Algoritma Genetik dengan melibatkan sebelas fitur teks. Fitur-fitur yang dijadikan acuan penentuan kalimat penting dalam penelitian tersebut adalah posisi kalimat, *positive keyword*, *negative keyword*, kemiripan antar kalimat, kalimat yang menyerupai judul, kalimat yang mengandung nama entiti, kalimat yang mengandung data numerik, koneksi antar kalimat, penjumlahan bobot antar kalimat, dan kalimat semantik. Percobaan ini memperoleh tingkat akurasi dengan pemampatan 30%, 20%, dan 10%, sebesar 47.46%, 41.29%, dan 35.01%.

Sebelas fitur teks yang digunakan dalam penelitian (Aristoteles et al., 2012) menjadi acuan pada penelitian yang dilakukan oleh (Wibowo et al., 2013) dengan metode Regresi Logistik Biner. Percobaan ini menghasilkan tingkat akurasi sebesar 91,1% dengan pemampatan 30%. Hasil uji parameter dalam percobaan ini menunjukkan fitur teks *"positive keyword"* memiliki pengaruh besar terhadap akurasi ringkasan.

Menurut (Raj et al., 2015) dalam penelitian *Sentence Boundary Disambiguation* dengan teks berbahasa Urdu menggunakan metode *Neural Network*, identifikasi batas kalimat merupakan faktor utama yang dijadikan acuan untuk menentukan

tingkat kepentingan teks. Setelah batas teks teridentifikasi, dihitung juga nilai ambang batas teks dan ukuran variasi. Dari penelitian ini diperoleh hasil akurasi 93.05% presisi, 99.53% recall, dan 96.18% f-measure.

Peringkasan teks dengan pendekatan ekstraktif dan menggunakan metode *Neural Network* diterapkan dalam penelitian yang dilakukan oleh (Kaikhah, 2004) yang menghasilkan tingkat akurasi rata-rata 96%. Metode ini menganalisa sebanyak tujuh fitur teks untuk menentukan kalimat penting dari dokumen. Fitur-fitur yang digunakan adalah kalimat yang mirip dengan judul, posisi kalimat, posisi paragraf dalam dokumen, kalimat pertama dalam paragraf, panjang kalimat, nomor tematik kata dalam kalimat, nomor kata yang serupa dengan judul di dalam kalimat.

Berdasarkan beberapa penelitian rujukan yang telah disebutkan, penelitian ini fokus pada pengujian hasil peringkasan teks menggunakan metode *feed forward neural network* (selanjutnya akan disingkat FFNN) dan kombinasi sebelas fitur teks yang diambil dari penelitian (Aristoteles et al., 2012).

1.2 Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah bagaimana tingkat akurasi peringkasan teks menggunakan metode FFNN dengan kombinasi sebelas fitur teks.

1.3 Batasan Masalah

Batasan masalah pada penelitian ini adalah :

1. Proses pengujian metode peringkasan teks FFNN dilakukan pada dokumen berbahasa Indonesia.
2. Tema dokumen yang digunakan adalah berita nasional.
3. Dokumen yang digunakan berjenis *plain* teks (berekstensi *.txt).

1.4 Tujuan

Tujuan dari penelitian ini adalah :

1. Mengimplementasikan metode FFNN pada sistem peringkasan teks.
2. Menganalisa fitur teks yang paling berpengaruh dalam proses peringkasan dokumen.
3. Mengevaluasi kinerja FFNN dibandingkan dengan Algoritma Genetika dan Regresi Logistik Biner berdasarkan hasil akurasi ringkasan teks yang diperoleh.

1.5 Manfaat

Manfaat dari penelitian ini adalah :

1. Memperoleh hasil ringkasan teks dalam waktu singkat dengan akurasi yang baik.
2. Mempermudah proses peringkasan teks untuk dokumen berskala besar.

II. TINJAUAN PUSTAKA

2.1 Peringkasan Teks

Peringkasan teks otomatis adalah proses mengurangi teks pada dokumen dengan menggunakan program komputer untuk membuat ringkasan yang berisikan poin-poin penting dimana hasil ringkasan tidak lebih dari setengah dokumen asli (Radev et al., 2002). Terdapat dua bagian dari kriteria peringkasan teks yaitu ekstraksi dan abstraksi. Teknik ekstraksi yaitu teknik peringkasan secara lengkap yang terdiri dari urutan-urutan kalimat yang disalin dan memilih bagian-bagian kalimat penting dari dokumen asli. Sedangkan teknik abstraksi adalah teknik peringkasan dengan mengambil informasi penting dari dokumen kemudian menghasilkan ringkasan yang menggunakan kalimat baru yang tidak terdapat pada dokumen asli (Suanmali et al., 2009).

2.2 Ekstraksi Fitur Teks

Penelitian ini menggunakan sebelas fitur teks. Sepuluh fitur teks dirujuk dari penelitian yang dilakukan (Fattah & Ren, 2008), yaitu posisi kalimat, *positive keyword*¸ *negative keyword*, kemiripan antar kalimat, kalimat yang menyerupai judul kalimat, kalimat yang mengandung nama entiti, kalimat yang mengandung data numerik, panjang kalimat, koneksi antar kalimat, dan penjumlahan bobot koneksi antar-kalimat. Fitur teks ke sebelas adalah kalimat semantik yang dirujuk dari penelitian (Aristoteles et al., 2012).

2.2.1 Posisi Kalimat (f_1)

Posisi kalimat adalah letak kalimat dalam sebuah paragraf. f_1 dalam penelitian ini diasumsikan bahwa kalimat pertama pada tiap paragraf adalah kalimat yang paling penting. Oleh karena itu, penelitian ini mengurutkan kalimat tersebut berdasarkan posisinya. Perhitungaan skor fitur teks posisi kalimat dapat dirumuskan sebagai berikut :

$$Score_{f_1}(s) = \frac{X}{N} \quad (1)$$

s : kalimat di dalam paragraf

f_1 : fitur teks posisi kalimat

N : total jumlah kalimat dalam satu paragraf

X : posisi kalimat dalam tiap paragraf

Penerapan fitur teks posisi kalimat diilustrasikan sebagai berikut :

Kebakaran di Pasar Sukaramai (6 kalimat)

Kebakaran yang terjadi di Pasar Sukaramai, Pekanbaru, Riau, Selasa (8/12) membuat panik para pedagang. Mereka berlarian keluar sambil berusaha menyelamatkan barang dagangannya yang masih tersisa. Peristiwa kebakaran ini menghanguskan puluhan kios yang berada di lantai dasar. Diduga api berasal dari salah satu kios di lantai dasar yang terletak di bagian tengah. Petugas sempat mengalami kesulitan dalam memadamkan api. Kepala Polresta Pekanbaru, Kombes Haryanto Watratan, mengatakan pihaknya masih menyelidiki penyebab kebakaran ini.

Perhitungan skor posisi kalimat dari teks di atas adalah $\frac{6}{6}$ untuk kalimat pertama, $\frac{5}{6}$ untuk kalimat kedua, $\frac{4}{6}$ untuk kalimat ketiga, $\frac{3}{6}$ untuk kalimat keempat, $\frac{2}{6}$ untuk kalimat kelima, dan $\frac{1}{6}$ untuk kalimat terakhir.

Kalimat pertama dalam setiap paragraf akan mendapatkan nilai yang paling tinggi sedangkan kalimat terakhir akan mendapatkan nilai yang paling kecil. Kalimat pertama umumnya merupakan ide pokok dari sebuah paragraf, sedangkan kalimat selanjutnya merupakan kalimat penjelas yang mendukung ide pokok paragraf.

2.2.2 *Positive Keyword* (f_2)

Positive keyword adalah kata yang paling banyak muncul pada sebuah kalimat. Perhitungan skor fitur teks *positive keyword* dapat dirumuskan sebagai berikut :

$$Score_{f_2}(s) = \frac{1}{length(s)} \sum_{i=1}^{n} tf_i * P(s \in S | keyword_i) \quad (2)$$

s : kalimat dalam ringkasan

S : kalimat dalam dokumen

f_2 : fitur teks *positive keyword*

n : jumlah *keyword* dalam kalimat

tf_i : adalah banyaknya *keyword* ke-i yang muncul dalam kalimat

$$P(s \in S | keyword_i) = \frac{P(keyword_i | s \in S) P(s \in S)}{P(keyword_i)}$$

$$P(keyword_i | s \in S) = \frac{(kalimat\ dalam\ ringkasan\ dan\ mengandung\ keyword_i)}{(kalimat\ dalam\ ringkasan)}$$

$$P(s \in S) = \frac{(kalimat\ dalam\ korpus\ pelatihan\ dan\ dalam\ ringkasan)}{(kalimat\ dalam\ korpus\ pelatihan)}$$

$$P(keyword_i) = \frac{(kalimat\ dalam\ korpus\ pelatihan\ dan\ mengandung\ keyword_i)}{(kalimat\ dalam\ korpus\ pelatihan)}$$

$P(s \in S|keyword_i)$ adalah perhitungan dari korpus pelatihan (ringkasan manual), tf_i, n, dan panjang kalimat dihitung menggunakan kalimat "*s*" pada tahap pengujian.

Penerapan fitur teks *positive keyword* diilustrasikan sebagai berikut :

Contoh teks :

Peristiwa **kebakaran** yang terjadi di Pasar Sukaramai, Pekanbaru, Riau, Selasa (8/12) membuat panik para pedagang. Mereka berlarian keluar sambil berusaha menyelamatkan barang dagangannya yang masih tersisa. Peristiwa **kebakaran** ini menghanguskan puluhan kios yang berada di lantai dasar. Diduga api berasal dari salah satu kios di lantai dasar yang terletak di bagian tengah. Petugas sempat mengalami kesulitan dalam memadamkan api. Kepala Polresta Pekanbaru, Kombes Haryanto Watratan, mengatakan pihaknya masih menyelidiki penyebab **kebakaran** ini.

Contoh ringkasan :

Peristiwa **kebakaran** yang terjadi di Pasar Sukaramai, Pekanbaru Riau, Selasa (8/12) membuat panik para pedagang. Peristiwa **kebakaran** ini menghanguskan puluhan kios yang berada di lantai dasar.

Contoh *positive keyword* yang diambil dari teks di atas adalah kata "**kebakaran**". Pada contoh teks berita, kata "**kebakaran**" terdapat dalam tiga kalimat dari keseluruhan dokumen yang berjumlah enam kalimat, sehingga diperoleh hasil

perhitungan $P(kebakaran)$ sebesar $\frac{3}{6}$. Kalimat ringkasan dari teks berita tersebut sebanyak dua kalimat yang keduanya mengandung kata "**kebakaran**", sehingga diperoleh $P(s \in S)$ sebesar $\frac{2}{6}$ dan $P(kebakaran|s \in S)$ sebesar $\frac{2}{2}$, $P(s \in S|kebakaran)$ sebesar $\frac{2}{3}$, dan skor kalimat pertama pada teks sebesar $\frac{1}{2}$.

2.2.3 *Negative Keyword* (f_3)

Negative keyword merupakan kebalikan dari fitur teks *positive keyword*. *Negative keyword* adalah kata yang sedikit muncul dalam kalimat. Perhitungan skor fitur *negative keyword* dapat dirumuskan sebagai berikut :

$$Score_{f_3}(s) = \frac{1}{length(s)} \sum_{i=1}^{n} tf_i * P(s \notin S|keyword_i) \quad (3)$$

s : kalimat dalam ringkasan

S : kalimat dalam dokumen

f_3 : fitur teks *negative keyword*

N : jumlah *keyword* dalam kalimat

tf_i : banyaknya *keyword* ke-i yang muncul dalam kalimat

Penerapan fitur teks *negative keyword* diilustrasikan sebagai berikut :

Contoh teks :

Peristiwa kebakaran yang terjadi di Pasar Sukaramai, Pekanbaru, Riau, Selasa (8/12) membuat panik para pedagang. Mereka berlarian keluar sambil berusaha menyelamatkan barang dagangannya yang masih tersisa. Peristiwa kebakaran ini menghanguskan puluhan kios yang berada di lantai dasar. Diduga api berasal dari salah satu kios di lantai dasar yang terletak di bagian tengah. **Petugas** sempat mengalami kesulitan dalam memadamkan api. Kepala

Polresta Pekanbaru, Kombes Haryanto Watratan, mengatakan pihaknya masih menyelidiki penyebab kebakaran ini.

Contoh ringkasan :

Peristiwa kebakaran yang terjadi di Pasar Sukaramai, Pekanbaru Riau, Selasa (8/12) membuat panik para pedagang. Peristiwa kebakaran ini menghanguskan puluhan kios yang berada di lantai dasar.

Contoh *negative keyword* yang diambil dari teks di atas adalah kata "**petugas**", sehingga diperoleh hasil perhitungan $P(petugas)$ sebesar $\frac{1}{6}$, $P(s \notin S)$ sebesar $\frac{4}{6}$, $P(petugas|s \notin S)$ sebesar $\frac{2}{2}$, $P(s \notin S|petugas)$ sebesar 4, dan skor kalimat pertama pada teks sebesar $\frac{6}{2}$.

2.2.4 Kemiripan Antar-Kalimat (f_4)

Kemiripan antar-kalimat merupakan kata yang muncul dalam kalimat sama dengan kata yang muncul dalam kalimat lain.

Perhitungan skor fitur kemiripan antar kalimat dapat dirumuskan sebagai berikut :

$$Score_{f_4}(s) = \left|\frac{Keyword\ dalam\ s\ \cap Keyword\ dalam\ antar\ kalimat}{Keyword\ dalam\ s\ \cup Keyword\ dalam\ antar\ kalimat}\right| \quad (4)$$

s : kalimat dalam dokumen

f_4 : fitur teks kemiripan antar kalimat

Contoh penerapan kemiripan antar-kalimat yaitu :

(1) Peristiwa kebakaran yang terjadi di Pasar Sukaramai, Pekanbaru, Riau, Selasa (8/12) membuat panik para pedagang.

(2) Peristiwa kebakaran ini menghanguskan puluhan kios yang berada di lantai dasar.

Perhitungan skor fitur teks kemiripan antar-kalimat dari contoh tersebut, diperoleh hasil skor kalimat pertama adalah $\frac{2}{18}$ dan skor kalimat kedua adalah $\frac{2}{18}$. Kalimat pertama memiliki dua kata yang sama dengan kalimat kedua yaitu "**peristiwa** dan **kebakaran**". Jumlah kata unik dari kalimat pertama dan kedua adalah 18 kata.

2.2.5 Kalimat yang Menyerupai Judul Dokumen (f_5)

Kalimat yang menyerupai judul dokumen adalah kata yang muncul dalam kalimat sama dengan kata yang ada dalam judul dokumen. Contoh penerapan kemiripan antar-kalimat yaitu :

Judul Berita : Kebakaran di Pasar Sukaramai

Kebakaran yang terjadi di Pasar Sukaramai, Pekanbaru, Riau, Selasa (8/12) membuat panik para pedagang. Mereka berlarian keluar sambil berusaha menyelamatkan barang dagangannya yang masih tersisa. Peristiwa kebakaran ini menghanguskan puluhan kios yang berada di lantai dasar.

Perhitungan skor fitur teks kalimat dari contoh tersebut, skor kalimat pertama yang menyerupai judul adalah $\frac{2}{11}$, skor kalimat kedua adalah $\frac{0}{12}$, dan skor kalimat ketiga adalah $\frac{1}{10}$. Kalimat pertama memiliki dua kata yang sama dengan judul dokumen yaitu kata "**kebakaran**" dan "**pasar sukaramai**". Kalimat kedua tidak

memiliki kata yang sama dengan judul dokumen. Kalimat ketiga memiliki satu kata yang sama dengan judul dokumen yaitu kata "**kebakaran**". Jumlah kata unik dalam judul ditambah kalimat pertama sebanyak 11 kata. Jumlah kata unik dalam judul ditambah kalimat kedua sebanyak 12 kata. Jumlah kata unik dalam judul ditambah kalimat ketiga sebanyak 10 kata.

Perhitungan skor fitur teks kalimat yang menyerupai judul dokumen dapat dihitung dengan menggunakan rumus berikut :

$$Score_{f_5}(s) = \left|\frac{Keyword\ dalam\ s\ \cap Keyword\ dalam\ judul}{Keyword\ dalam\ s\ \cup Keyword\ dalam\ judul}\right| \quad (5)$$

s : kalimat dalam dokumen

f_5 : fitur teks kalimat yang menyerupai judul dokumen

2.2.6 Kalimat yang Mengandung Nama Entiti (f_6)

Nama entiti adalah sebuah kumpulan kata yang memiliki makna atau membentuk nama sebuah institusi, nama orang, nama pulau. Misalnya Institut Pertanian Bogor merupakan kumpulan kata yang memiliki makna sebuah institusi perguruan tinggi.

Skor fitur kalimat yang mengandung nama entiti dapat diperoleh dengan rumus perhitungan sebagai berikut :

$$Score_{f_6}(s) = \frac{nama\ entiti\ dalam\ (s)}{Panjang\ kalimat\ (s)} \quad (6)$$

s : kalimat dalam dokumen

f_6 : fitur teks kalimat yang mengandung nama entiti

Penerapan fitur teks keenam diilustrasikan sebagai berikut :

> Peristiwa kebakaran yang terjadi di **Pasar Sukaramai**, Pekanbaru, Riau, Selasa (8/12) membuat panik para pedagang. Mereka berlarian keluar sambil berusaha menyelamatkan barang dagangannya yang masih tersisa. Peristiwa kebakaran ini menghanguskan puluhan kios yang berada di lantai dasar. Diduga api berasal dari salah satu kios di lantai dasar yang terletak di bagian tengah. Petugas sempat mengalami kesulitan dalam memadamkan api. Kepala **Polresta Pekanbaru**, Kombes **Haryanto Watratan**, mengatakan pihaknya masih menyelidiki penyebab kebakaran ini.

Perhitungan skor fitur ke enam berdasarkan contoh teks di atas adalah $\frac{1}{12}$ untuk kalimat pertama. Kalimat pertama memiliki satu nama entiti dari 12 kata unik, yaitu "**Pasar Sukaramai**". Kalimat terakhir mengandung dua nama entiti dari 10 kata unik yaitu "**Polresta Pekanbaru**" dan "**Haryanto Watratan**", sehingga skornya adalah $\frac{2}{10}$. Kalimat lainnya tidak memiliki nama entity, sehingga diberi nilai 0.

2.2.7 Kalimat yang Mengandung Data Numerik (f_7)

Data numerik merupakan salah satu fitur yang dianggap penting dalam peringkasan teks karena dalam kalimat yang berisi data numerik terdapat kalimat yang penting. Dokumen keuangan merupakan contoh dokumen yang banyak mengandung data numerik, sehingga peringkasan dokumen tersebut mempertimbangkan fitur teks data numerik. Perhitungan fitur teks kalimat mengandung data numerik dapat diperoleh dengan rumus berikut :

$$Score_{f_7}(s) = \frac{data\ numerik\ dalam\ (s)}{Panjang\ kalimat\ (s)} \quad (7)$$

s : kalimat dalam dokumen

f_7 : fitur teks kalimat yang mengandung data numerik

Penerapan fitur teks ketujuh diilustrasikan sebagai berikut :

> Peristiwa kebakaran yang terjadi di Pasar Sukaramai, Pekanbaru, Riau, Selasa **(8/12)** membuat panik para pedagang. Mereka berlarian keluar sambil berusaha menyelamatkan barang dagangannya yang masih tersisa. Peristiwa kebakaran ini menghanguskan puluhan kios yang berada di lantai dasar. Diduga api berasal dari salah satu kios di lantai dasar yang terletak di bagian tengah. Petugas sempat mengalami kesulitan dalam memadamkan api. Kepala Polresta Pekanbaru, Kombes Haryanto Watratan, mengatakan pihaknya masih menyelidiki penyebab kebakaran ini.

Perhitungan skor fitur ke tujuh berdasarkan contoh teks di atas adalah $\frac{1}{12}$ untuk kalimat pertama. Kalimat pertama memiliki satu data numerik dari 12 kata unik, yaitu "**(8/12)**". Sedangkan kalimat lainnya tidak memiliki data numerik, sehingga skornya adalah 0.

2.2.8 Panjang Kalimat (f_8)

Panjang kalimat dihitung berdasarkan jumlah kata dalam kalimat dibagi jumlah kata unik dalam dokumen. Perhitungan skor fitur teks panjang kalimat dapat dirumuskan sebagai berikut :

$$Score_{f_8}(s) = \frac{jumlah\ kata\ dalam\ (s)}{kata\ unik\ dalam\ dokumen} \quad (8)$$

s : kalimat dalam dokumen

f_8 : fitur teks panjang kalimat

Penerapan fitur teks panjang kalimat diilustrasikan sebagai berikut :

> Peristiwa kebakaran yang terjadi di Pasar Sukaramai, Pekanbaru, Riau, Selasa (8/12) membuat panik para pedagang. Mereka berlarian keluar sambil berusaha menyelamatkan barang dagangannya yang masih tersisa. Peristiwa kebakaran ini menghanguskan puluhan kios yang berada di lantai dasar. Diduga api berasal dari salah satu kios di lantai dasar yang terletak di bagian tengah. Petugas sempat mengalami kesulitan dalam memadamkan api. Kepala Polresta Pekanbaru, Kombes Haryanto Watratan, mengatakan pihaknya masih menyelidiki penyebab kebakaran ini.

Langkah pertama untuk mendapatkan skor fitur panjang kalimat adalah menghitung jumlah kata unik yang terdapat dalam dokumen. Jumlah kata unik pada contoh teks berita adalah 57 kata. Jumlah kata tersebut akan dibandingkan dengan jumlah kata dari setiap kalimat. Jumlah kata unik pada kalimat pertama pada teks adalah 11 kata, kalimat kedua sebanyak 9 kata, kalimat ketiga sebanyak 8 kata, kalimat ke empat sebanyak 11 kata, kalimat ke lima sebanyak 6 kata, dan kalimat terakhir sebanyak 12 kata. Jumlah tersebut menghasilkan skor fitur panjang kalimat sebesar $\frac{11}{57}$ untuk kalimat pertama, $\frac{9}{57}$ untuk kalimat kedua, $\frac{8}{57}$

untuk kalimat ketiga, $\frac{11}{57}$ untuk kalimat keempat, $\frac{6}{57}$ untuk kalimat kelima, dan $\frac{12}{57}$ untuk kalimat keenam.

2.2.9 Koneksi Antar-Kalimat (f_9)

Koneksi antar-kalimat adalah banyaknya kalimat yang memiliki kata yang sama dengan kalimat lain dalam satu dokumen. Perhitungan fitur teks koneksi antar-kalimat dapat menggunakan rumus perhitungan berikut :

$$Score_{f_9}(s) = \#Jumlah\ koneksi\ antar-kalimat \quad (9)$$

s : kalimat dalam dokumen

f_9 : fitur teks jumlah koneksi antar-kalimat

Penerapan fitur teks koneksi antar-kalimat diilustrasikan sebagai berikut :

> **Peristiwa kebakaran** yang terjadi di Pasar Sukaramai, **Pekanbaru**, Riau, Selasa (8/12) membuat panik para pedagang. Mereka berlarian keluar sambil berusaha menyelamatkan barang dagangannya yang **masih** tersisa. **Peristiwa kebakaran** ini menghanguskan puluhan **kios** yang berada di **lantai dasar**. Diduga **api** berasal dari salah satu **kios** di **lantai dasar** yang terletak di bagian tengah. Petugas sempat mengalami kesulitan dalam memadamkan **api**. Kepala Polresta **Pekanbaru**, Kombes Haryanto Watratan, mengatakan pihaknya **masih** menyelidiki penyebab **kebakaran** ini.

Perhitungan skor fitur koneksi antar-kalimat berdasarkan contoh teks di atas adalah 3 untuk kalimat pertama, yaitu kata **peristiwa**, **kebakaran**, dan **Pekanbaru**. Kalimat kedua memiliki skor 1, yaitu kata **masih**. Kalimat ketiga memiliki skor 5, yaitu kata **peristiwa**, **kebakaran**, **kios**, **lanta**i, dan **dasar**.

Kalimat keempat memiliki skor 4, yaitu kata **api**, **kios**, **lantai**, dan **dasar**. Kalimat kelima memiliki skor 1 karena hanya terkoneksi melalui kata **api**. Kalimat keenam memiliki skor 3, yaitu kata **pekanbaru**, **masih**, dan **kebakaran**.

2.2.10 Penjumlahan Bobot Koneksi Antar-Kalimat (f_{10})

Fungsi fitur teks ini adalah menjumlahkan bobot koneksi antar-kalimat. Perhitungan fitur teks penjumlahan bobot koneksi antar-kalimat dapat meggunakan rumus perhitungan berikut :

$$Score_{f_{10}}(s) = \sum koneksi\ antar\ kalimat \tag{10}$$

s : kalimat dalam dokumen

f_{10} : fitur teks penjumlahan bobot koneksi antar-kalimat

2.2.11 Kalimat Semantik (f_{11})

Kalimat semantik adalah kalimat yang mencirikan hubungan antar kalimat yang didasari semantik. Asumsikan D adalah sebuah dokumen, $t(|t| = M)$ adalah banyaknya kata dalam D, dan $S(|S| = N)$ adalah banyaknya kalimat dalam D. Matriks kata dapat dilihat pada Rumus (11), dengan S_j adalah kalimat ke-j dalam dokumen dan t_i adalah *term* ke-i yang muncul didalam dokumen. Penelitian ini menggunakan semua *keyword* atau *term* yang ada dalam dokumen kecuali kata-kata *stoplist*.

$$A = \begin{array}{c|ccccc} & S_1 & S_2 & \dots & S_N \\ \hline t_1 & w_{1,1} & w_{1,2} & \dots & w_{1,N} \\ t_2 & w_{2,1} & w_{2,2} & \dots & wa_{2,N} \\ \vdots & \vdots & \vdots & \ddots & \dots \\ t_M & w_{M,1} & w_{2,1} & \dots & w_{M,N} \end{array} \tag{11}$$

dengan $w_{i,j}$ didefinisikan pada Rumus (12), dan tf_i adalah banyaknya kemunculan *term* ke-i pada kalimat. SF_i s*entences frequency*$_i$ merupakan banyak kalimat yang mengandung *term* ke-i, sedangkan $ISF_i = log\left(\frac{N}{SF_i}\right)$ merupakan ukuran diskriminan kemunculan *term* ke-*i* dalam dokumen, *N* adalah banyaknya kalimat dalam satu dokumen.

$$w_{i,j} = tf_i \times ISF_i \tag{12}$$

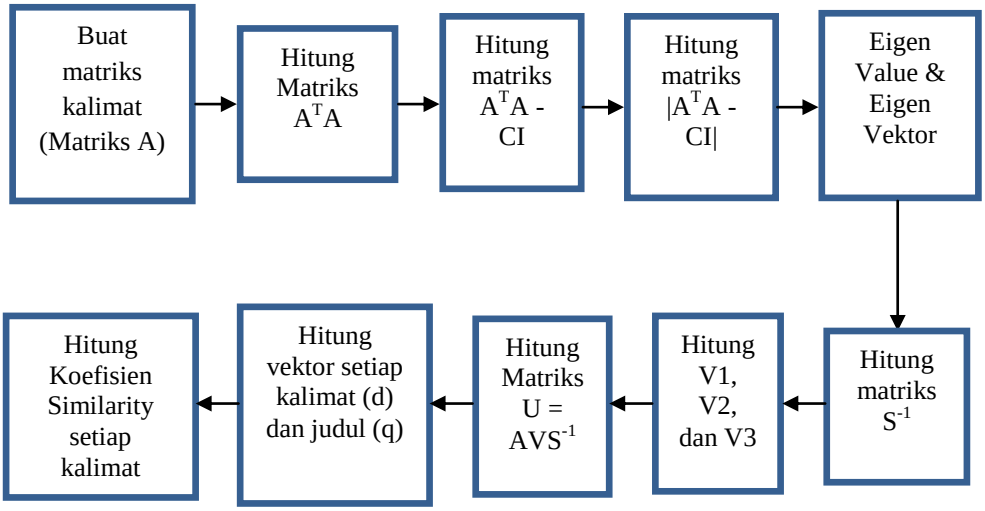
Kalimat semantik, dalam penelitian ini, ditentukan dengan menggunakan teknik SVD (Yeh et al., 2005). Persamaan *Singular Value Decomposition* (SVD) adalah $A = USV^T$, dengan U adalah matriks $M \times M$ vektor *singular* kiri, S adalah matriks diagonal $M \times N$ *singular value*, dan V adalah matriks $N \times N$ vektor *singular* kanan. Dimana vektor *V* merepresentasikan kalimat, sedangkan vektor *U* merepresentasikan kata yang ada pada suatu dokumen. Vektor S merupakan tingkat penciri dari matrik *A*. Skor fitur teks kalimat semantik dapat diterapkan dengan rumus berikut :

$$Score_{f_{11}}(s) = matriks\ kemiripan\ (s) \tag{13}$$

s : kalimat dalam dokumen

f_{11} : fitur teks kalimat semantik

Perhitungan kalimat semantik diilustrasikan sebagai berikut :



Gambar 2. 1 Ilustrasi perhitungan skor kalimat semantik

Judul : Input Jaringan Syaraf

(1) Jaringan menerima input.

(2) Input memiliki koneksi berbobot.

(3) Sel syaraf memiliki ambang input.

$$\text{Matriks } A = \begin{matrix} d_1 & d_2 & d_3 \end{matrix} \quad Q$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

a. Menghitung matriks A^TA

$$\text{Matriks } A^TA = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 4 & 2 \\ 1 & 2 & 5 \end{bmatrix}$$

b. Menghitung matriks A^TA - CI

$$A^TA - CI = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 4 & 2 \\ 1 & 2 & 5 \end{bmatrix} - \begin{bmatrix} c & 0 & 0 \\ 0 & c & 0 \\ 0 & 0 & c \end{bmatrix} = \begin{bmatrix} 3-c & 1 & 1 \\ 1 & 4-c & 2 \\ 1 & 2 & 5-c \end{bmatrix}$$

c. Menghitung $|A^TA - CI|$

$$|A^TA - CI| = -c^3 + 12c^2 - 41c + 43 = 0$$

d. Mencari Eigen Value & Eigen Vektor

Eigenvalue = *C1* : 7,049; *C2* : 2,308; *C3* : 2,643

Eigenvektor = *V1* : 0.56, 1, 1.25; *V2* : -0.79, 1, -0.45; *V3* : 2.24, 1, -1.80

e. Mencari matriks S^{-1}

$$\text{Matriks } S = \begin{bmatrix} \sqrt{c1} & 0 & 0 \\ 0 & \sqrt{c2} & 0 \\ 0 & 0 & \sqrt{c3} \end{bmatrix} \Rightarrow \begin{bmatrix} 2.65 & 0 & 0 \\ 0 & 1.52 & 0 \\ 0 & 0 & 1.63 \end{bmatrix}$$

$$\text{Matriks } S^{-1} = \begin{bmatrix} 0.38 & 0 & 0 \\ 0 & 0.66 & 0 \\ 0 & 0 & 0.61 \end{bmatrix}$$

f. Menghitung vektor dari *V1*, *V2*, dan *V3*

$|V1| = \sqrt{(0.56)^2 + (1)^2 + (1.25)^2} = 1.696$

$|V2| = \sqrt{(-0.79)^2 + (1)^2 + (-0.45)^2} = 1.35$

$|V3| = \sqrt{(2.24)^2 + (1)^2 + (-1.8)^2} = 3.04$

$$\text{Matriks } V = \begin{bmatrix} 0.33 & 0.59 & 0.74 \\ -0.58 & 0.74 & -0.33 \\ 0.74 & 0.33 & -0.59 \end{bmatrix}$$

g. Mencari Matriks $U = AVS^{-1}$

$$\text{Matriks } U = \begin{bmatrix} 0.1254 & 0.3894 & 0.4514 \\ 0.1254 & 0.3894 & 0.4514 \\ 0.1862 & 1.0956 & -0.1098 \\ 0.0608 & 0.7062 & -0.5612 \\ -0.2204 & 0.5148 & -0.2013 \\ -0.2204 & 0.5148 & -0.2013 \\ 0.2812 & 0.2178 & -0.3599 \\ 0.2812 & 0.2178 & -0.3599 \\ 0.2812 & 0.2178 & -0.3599 \end{bmatrix}$$

h. Menghitung vektor setiap kalimat *(d)* dan judul *(q)*

Vektor $d = d^T US^{-1}$	Vektor $q = q^T US^{-1}$

d_1 = [0.167 1.237 0.483]
d_2 = [-0.074 1.869 -0.665]
d_3 = [0.415 1.621 0.251]
q = [0.226 1.124 0.428]

i. Menghitung Koefisien Similarity setiap kalimat *(sim(q,d))*

$$sim(q, d_1) = \frac{0.038 + 1.39 + 0.207}{\sqrt{1.497}\sqrt{1.791}} = 0.997$$

$$sim(q, d_2) = \frac{0.017 + 2.1 - 0.285}{\sqrt{1.497}\sqrt{3.94}} = 0.74$$

$$sim(q, d_3) = \frac{0.094 + 1.822 - 0.107}{\sqrt{1.497}\sqrt{2.863}} = 0.977$$

Berdasarkan perhitungan di atas, kalimat pertama memiliki skor 0.997, kalimat kedua memiliki skor 0.74, dan kalimat ketiga memiliki nilai 0.977.

2.3 Sejarah Jaringan Syaraf Tiruan

Menurut (Siang, 2005), Jaringan Syaraf Tiruan (JST) diperkenalkan pertama kali pada tahun 1943 oleh *neurophysiologist* Waren McCulloch dan *logician* Walter

Pits, namun teknologi yang tersedia pada saat itu belum memungkinkan mereka berbuat lebih jauh. JST didefinisikan sebagai susunan elemen-elemen penghitung yang disebut sebagai neuron atau titik (node) yang saling terhubung guna dimodelkan untuk meniru fungsi otak manusia. JST dicirikan dengan adanya proses pembelajaran yang berfungsi untuk mengadaptasi parameter-parameter jaringannya. Bobot dalam jaringan yang diusulkan oleh McCulloch dan Pitts diatur untuk melakukan fungsi logika sederhana. Fungsi aktivasi yang dipakai adalah fungsi *threshold*.

Tahun 1958, Rossenbaltt memperkenalkan dan mulai mengembangkan model jaringan yang disebut *perceptron*. Metode pelatihan diperkenalkan untuk mengoptimalkan hasil iterasinya. Widrow dan Holf (1960) mengembangkan *perceptron* dengan memperkenalkan aturan pelatihan jaringan, yang dikenal sebagai aturan *Delta* (atau sering disebut kuadrat rata-rata terkecil). Aturan ini akan mengubah bobot *perceptron* apabila keluaran yang dihasilkan tidak sesuai dengan targrt yang diinginkan. Apa yang dilakukan peneliti terdahulu hanya menggunakan jaringan dengan layer tunggal (*single layer*). JST ditentukan oleh beberapa hal berikut :

1. Pola hubungan antar neuron (disebut arsitektur jaringan)
2. Metode untuk menentukan bobot penghubung (disebut pembelajaran)
3. Penggunaan bias
4. Fungsi aktivasi

2.3.1 Dasar Jaringan Syaraf Tiruan

JST adalah proses paradigma suatu informasi yang terinspirasi oleh sistem syaraf biologi, sama seperti otak yang memproses suatu informasi. Elemen mendasar dari paradigma tersebut adalah struktur yang baru dari sistem proses informasi. JST, seperti manusia, belajar dari suatu contoh. JST dibentuk untuk memecahkan masalah tertentu seperti pengenalan pola atau klasifikasi karena proses pembelajaran.

JST dibentuk sebagai generalisasi model Matematika dari syaraf Biologi, dengan asumsi bahwa :

a. Proses informasi terjadi pada banyak elemen sederhana (neuron).
b. Sinyal dikirim diantara neuron-neuron melalui penghubung-penghubung.
c. Penghubung, antar neuton memiliki bobot yang akan memperkuat atau memperlemah sinyal.
d. Setiap neuron menggunakan fungsi aktivasi yang dikenakan pada jumlahan input yang diterima untuk menentukan output. Biasanya output ini selanjutnya dibandingkan dengan suatu ambang batas (Siang, 2005).

2.3.2 Model Dasar Jaringan Syaraf Tiruan

JST mengadopsi esensi dasar dari sistem syaraf Biologi, sehingga syaraf tiruan digambarkan sebagai berikut : menerima masukan baik dari data yang dimasukkan atau dari output sel syaraf pada jaringan syaraf. Setiap input datang melalui suatu koneksi atau hubungan yang mempunyai sebuah bobot (*weight*). Setiap sel syaraf mempunyai sebuah nilai ambang. Jumlah bobot dari input dan dikurangi dengan nilai ambang kemudian akan mendapatkan suatu aktivasi dari

sel syaraf (*post synaptic potential*, PSP, dari sel syaraf). Sinyal aktivasi kemudian menjadi fungsi aktivasi atau fungsi transfer untuk menghasilkan output dari sel syaraf. Jika tahapan fungsi aktivasi digunakan (output sel syaraf = 0 jika input < 0 dan 1 jika input >= 0) makat tindakan sel syaraf sama dengan sel syaraf Biologi yang dijelaskan di atas (pengurangan nilai ambang dari jumlah bobot dan membandingkan dengan 0 adalah sama dengan membandingkan jumlah bobot dengan nilai ambang). Biasanya tahapan fungsi jarang digunakan dalam JST.

2.3.3 Feedforward Neural Network (FFNN)

JST model ini merupakan tipe jaringan syaraf tiruan paling sederhana. Seluruh koneksi dari output suatu layer FFNN akan langsung menuju input layer berikutnya. Jaringan model ini juga tidak memiliki koneksi di dalam layer yang sama atau koneksi dari sebuah layer ke layer sebelumnya (Budiharto, 2016).

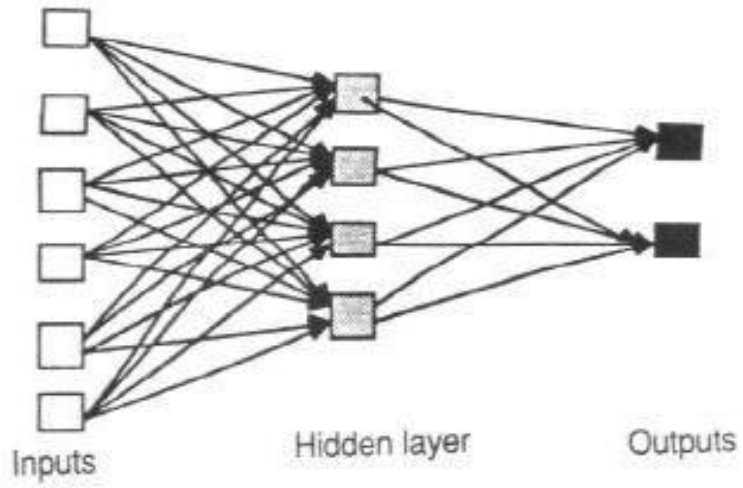
2.3.4 Arsitektur Jaringan

Arsitektur (pola koneksi) JST dibagi ke dalam dua kategori :

1. Struktur *feedforward*

   Sebuah jaringan yang sederhana mempunyai struktur *feedforward* (FF) dimana sinyal bergerak dari input kemudian melewati lapisan tersembunyi dan akhirnya mencapai unit output (mempunyai struktur perilaku yang stabil). Tipe jaringan FF mempunyai sel syaraf yang tersusun dari beberapa lapisan. Lapisan input bukan merupakan sel syaraf. Lapisan ini hanya member pelayanan dengan mengenalkan suatu nilai dari suatu variabel. Lapisan tersembunyi dan lapisan output sel syaraf terhubung satu sama lain dengan

lapisan sebelumnya. Kemungkinan yang timbul adalah adanya hubungan dengan beberapa unit dari lapisan sebelumnya atau terhubung semuanya (Siang, 2005).



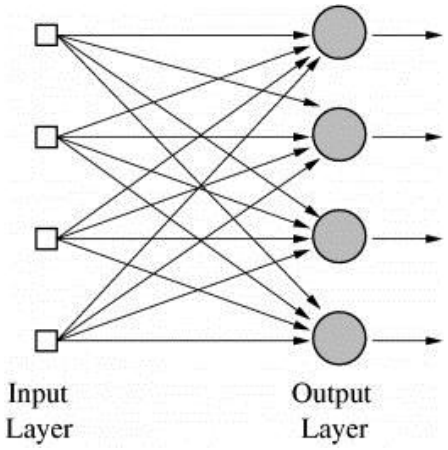
Gambar 2. 2 Jaringan Syaraf Tiruan *FeedForward*

Struktur FF adalah sebagai berikut:

- *Single-layer perceptron*
- *Multilayer perceptron*
- *Radial-basis function networks*
- *Higher-order networks*
- *Polynomial learning networks*

a. Jaringan Layer Tunggal (*Single Layer Network*)

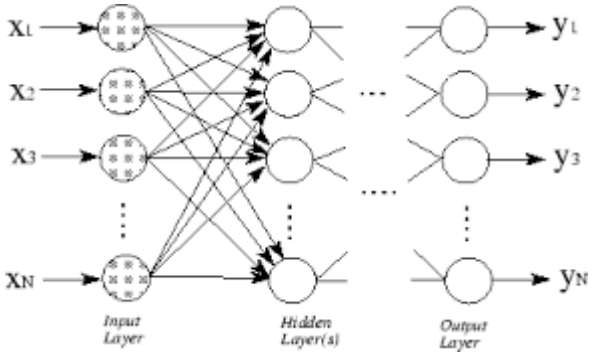
Jaringan ini dapat dilihat pada Gambar 2.3, sekumpulan input neuron dihubungkan langsung dengan sekumpulan outputnya. Beberapa model (misal *perceptron*), hanya ada sebuah unit neuron output.



Gambar 2. 3 *Single Layer Neural Network*

b. Jaringan Layer Jamak (*Multilayer Network*)

Jaringan layer jamak merupakan perluasan dari layer tunggal. Jaringan ini, selain unit input dan output, memiliki unit-unit lain (sering disebut layer tersembunyi atau *hidden layer*). Beberapa layer tersembunyi juga dapat menjadi penyusun jaringan ini. Sama seperti pada unit input dan output, unit-unit dalam satu layar tidak saling berhubungan.



Gambar 2. 4 *Multilayer Network*

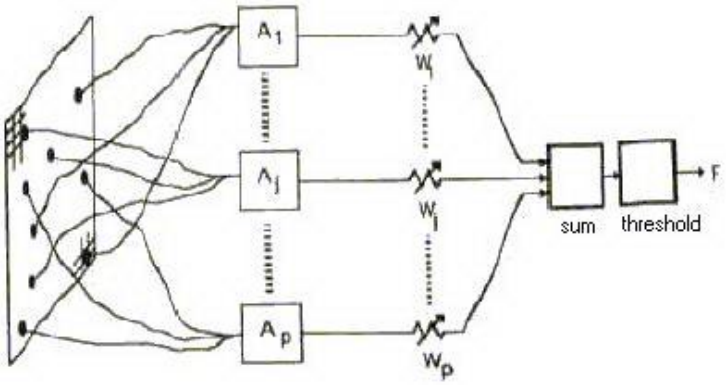
2.3.5 Neuron

Neuron dianalogikan dengan neurosikologi *(neurophysiology)* pada otak manusia. Dalam JST, neuron diartikan sebagai suatu bagian terkecil dari JST yang

berfungsi sebagai elemen proses. Neuron dapat dinyatakan sebagai *processor* sederhana dari sistem JST.

2.3.6 Perceptron

Perceptron termasuk ke dalam salah satu bentuk JST yang sederhana. Perceptron biasanya digunakan untuk mengklasifikasikan tipe pola tertentu yang sering dikenal dengan istilah pemisahan secara linear. Perceptron pada jaringan syaraf tiruan dengan satu lapis umumnya memiliki bobot yang bisa diatur dan suatu nilai ambang. Algoritma yang digunakan oleh aturan perceptron ini akan mengatur parameter-parameter bebasnya melalui proses pembelajaran. Perceptron memiliki tiga layer, yaitu unit sensor, unit assosiator, dan unit respon, yang menyerupai model retina. Perceptron dapat dilihat pada Gambar 2.5 (Budiharto, 2016).



Gambar 2. 5 Bentuk Peceptron

2.3.7 Paradigma Pembelajaran

Paradigma pembelajaran pada JST berdasarkan cara memodifikasi bobotnya yaitu pelatihan dengan *supervise* dan tanpa *supervise*. Pelatihan dengan *supervise* memiliki sejumlah pasang data yang dipakai untuk melatih jaringan hingga bobot yang dinginkan. Pasangan data tersebut berfungsi sebagai “guru” untut melatih

jaringan hingga diperoleh bentuk yang terbaik. “Guru” akan memberikan informasi yang jelas tentang bagaimana sistem harus mengubah dirinya untuk meningkatkan hasil kerjanya. Setiap kali pelatihan, suatu input diberikan ke jaringan. Jaringan akan memproses dan mengeluarkan keluaran. Selisih antara keluaran jaringan dengan keluaran yang diinginkan merupakan kesalahan yang terjadi. Jaringan akan memodifikasi bobot sesuai dengan kesalahan tersebut. Jaringan perceptron, adaline, dan *backpropagation* merupakan model-model yang menggunakan pelatihan dengan *supervise*. Sebaliknya, dalam pelatihan tanpa *supervise* tidak ada “guru” yang akan mengarahkan proses pelatihan.

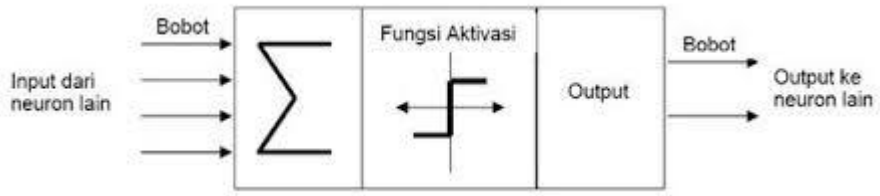
Perubahan bobot jaringan dilakukan berdasarkan parameter tertentu dan jaringan dimodifikasi menurut ukuran parameter tersebut dalam proses pelatihan. Sebagai contoh, dalam model jaringan kompetitif, jaringan terdiri dari dua layer, yaitu layer input dan kompetisi. Layer input menerima data eksternal. Layer kompetisi berisi neuron-neuron yang saling berkompetisi agar memperoleh kesempatan untuk merespon sifat-sifat yang ada dalam data masukan. Neuron yang memenangkan kompetisi akan memperoleh sinyal yang berikutnya akan diteruskan. Bobot neuron akan dimodifikasi sehingga lebih menyerupai data masukan (Siang, 2005).

2.3.8 Fungsi Aktivasi

Jaringan syaraf merupakan salah satu representasi buatan dari otak manusia yang selalu mencoba untuk mensimulasikan proses pembelajaran pada otak manusia. Istilah buatan digunakan karena jaringan syaraf ini diimplementasikan

menggunakan program komputer yang mampu menyelesaikan sejumlah proses perhitungan selama proses pembelajaran. Seperti halnya otak manusia, JST juga terdiri dari beberapa neuron yang saling berhubungan.

Neuron-neuron tersebut akan mentransformasikan informasi yang diterima melalui sambungan keluarnya menuju neuron yang lain. Hubungan ini dikenal dengan nama bobot. Informasi tersebut disimpan pada suatu nilai tertentu pada bobot tersebut (Siang, 2005).



Gambar 2. 6 Fungsi Aktivasi

Neuron buatan bekerja dengan cara yang sama dengan neuron Biologi. Sesuai dengan Gambar 2.6, informasi yang diterima (input) akan dikirim ke neuron dengan bobot kedatangan tertentu. Input ini akan diproses oleh suatu fungsi perambatan yang akan menjumlahkan nilai-nilai semua bobot yang datang. Hasil penjumlahan ini akan dibandingkan dengan suatu nilai ambang tertentu melalui fungsi aktivasi setiap neuron. Apabila input melewati suatu nilai ambang tertentu, maka neuron akan diaktifkan, dan sebaliknya. Jika neuron diaktifkan, maka neuron akan akan mengirim output melalui bobot-bobot outputnya ke semua neuron yang berhubungan dengannya, dan begitu seterusnya. Neuron akan dikumpulkan dalam lapisan yang disebut dengan lapisan neuron. Neuron-neuron pada satu lapisan akan dihubungkan dengan lapisan sebelumnya dan sesudahnya (kecuali lapisan input dan output). Informasi yang diberikan pada JST akan

dirambatkan dari satu lapisan ke lapisan lainnya. Ada beberapa fungsi aktivasi yang sering digunakan dalam JST, antara lain :

1. Fungsi Linear (*Purelin*)

   Fungsi linear memiliki nilai output yang sama dengan nilai inputnya.

2. Fungsi Bipolar (*Symentric Hardlims*)

   Fungsi bipolar sebenarnya hampir sama dengan fungsi undak biner, hanya saja output yang dihasilkan berupa 1 atau -1.

3. Fungsi Undak Biner (*Hardlimit*)

   Jaringan dengan lapis tunggal sering menggunakan fungsi undak biner untuk mengkonversikan input dari suatu variabel yang bernilai kontinu ke suatu output biner (0 atau 1).

4. Fungsi Saturating Linear (*Satlin*)

   Fungsi ini akan bernilai 0 jika inputnya kurang dari -1/2 dan akan bernilai 1 jika inputnya lebih dari 1/2. Sedangkan jika nilai input terletak antara -1/2 dan 1/2, maka outputnya akan bernilai sama dengan nilai input ditambah 1/2.

5. Fungsi Sigmoid Biner (*Logsig*)

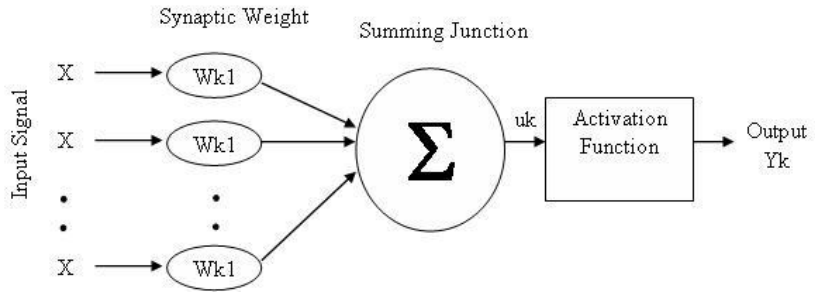
   Fungsi ini digunakan untuk jaringan syaraf yang dilatih dengan menggunakan metode *Backpropagation*. Fungsi Sigmoid Biner memiliki nilai pada *range* 0 sampai 1. Oleh karena itu, fungsi ini sering digunakan untuk jaringan syaraf yang membutuhkan nilai output yang terletak pada interval 0 sampai 1.

6. Fungsi Sigmoid Bipolar (*Tansig*)

   Fungsi ini hampir sama denga fungsi Sigmoid Biner, hanya saja output dari fungsi ini memiliki range antara 1 sampai -1 (Siang, 2005).

2.3.9 Bias

Lapisan neuron satu dengan yang lain dihubungkan dengan model koneksi yang memiliki bobot-bobot *w* dan *v*. Lapis tersembunyi dapat memiliki bias dengan bobot sama dengan satu. Berikut merupakan gambar JST dengan bias.



Gambar 2. 7 Jaringan Syaraf Tiruan dengan Bias

Neuron merupakan pengolahan informasi dasar dari suatu neural network yang terdiri dari :

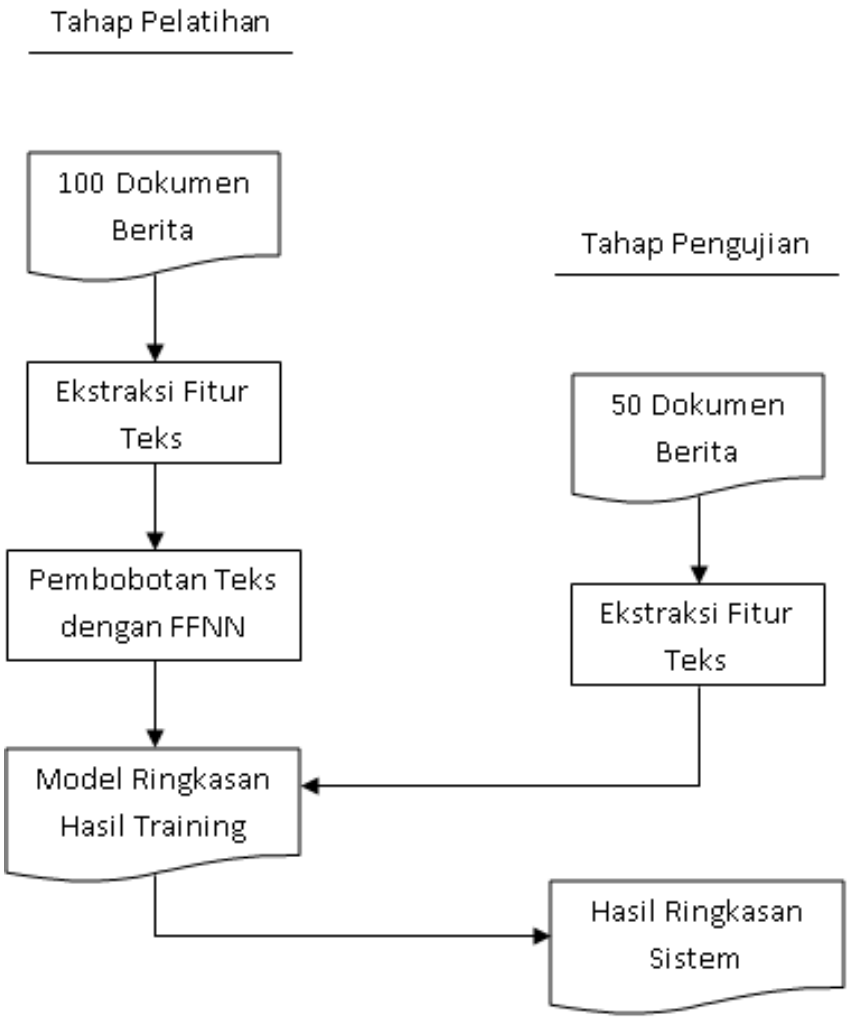
1. Himpunan jaringan (*links*), sebagai input dengan bobot.
2. Sebuah fungsi tambahan (*linear combiner*) untuk menghitung jumlah bobot dari semua input.
3. Sebuah fungsi aktivasi (*squashing function*) untuk membatasi amplitude dari output.

Bias merupakan parameter eksternal dari neuron. Hal ini dapat dimodelkan dengan penambahan ekstra input seperti pada Gambar 2.7 (Siang, 2005).

III. METODE PENELITIAN

3.1 Prosedur Penelitian

Penelitian ini dilakukan dengan tiga tahap, yaitu pengumpulan dokumen teks, proses *training* data (pelatihan), dan proses *testing* data (pengujian). Diagram alir tahap pelatihan dan pengujian data dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Diagram Alir Tahap Pelatihan dan Pengujian Peringkasan Teks

3.1.1 Tahap Pengumpulan Dokumen

Penelitian ini membutuhkan masukan berupa dokumen teks bahasa Indonesia berjenis file teks (ekstensi .txt). Tahap pelatihan pada penelitian ini menggunakan sebanyak 100 dokumen berita nasional yang belum dan telah diringkas. Tahap pengujian pada penelitian ini mengggunakan sebanyak 50 dokumen berita nasional yang belum dan telah diringkas. Dokumen yang digunakan diadopsi dari berita *online* harian Kompas dalam korpus (Ridha, 2002).

3.1.2 Tahap Pelatihan

Tahap pelatihan pada sistem ini melibatkan 4 jenis komputer dengan rincian spesifikasi sebagai berikut :

Spec (1) : Intel Core i5-7200U 4CPU @2.5-2.7 GHz RAM 4 GB

Spec (2) : Intel Core i5-3470S 3CPU @2.9 GHz RAM 1.5 GB

Spec (3) : Intel Core i3-3120M 4CPU @2.5 GHz RAM 2 GB

Spec (4) : AMD A10-5757M APU 4CPU @2.5 GHz RAM 4 GB

Tahap ini terdiri dari dua bagian, yaitu ekstraksi fitur teks dan pemodelan bobot fitur dengan algoritma FFNN. Penelitian ini tidak lagi melakukan ekstraksi fitur teks, namun menggunakan hasil ekstraksi fitur pada penelitian Aristoteles *et al* (2012). Fitur teks merupakan suatu proses perhitungan fitur-fittur untuk mendapatkan bobot sehingga dapat diketahui kalimat yang masuk dalam ringkasan dan tidak masuk ringkasan. Tahap ini melibatkan komputer dengan spesifikasi (1) sebanyak 3 komputer dan spesifikasi (2) sebanyak 7 komputer. Setiap komputer dilakukan percobaan sebanyak 5 kali untuk setiap variasi

pemampatan, sehingga total percobaan yang dilakukan adalah 150 kali untuk tiga variasi pemampatan.

Rata-rata waktu yang dibutuhkan pada tahap ini juga diukur. Perbedaan waktu diukur berdasarkan perbedaan spesifikasi komputer (1), (2), (3), dan (4). Setiap komputer dilakukan percobaan sebanyak 5 kali dengan pemampatan 10%. Total percobaan yang dilakukan sebanyak 20 kali.

3.1.3 Tahap Pengujian

Tahap pengujian digunakan 50 dokumen teks bahasa Indonesia (dokumen yang digunakan berbeda dengan tahap pelatihan). Peringkasan pada tahap pengujian didasari model yang telah terbentuk di tahap pelatihan. Hasil ringkasan sistem dibandingkan dengan ringkasan secara manual untuk mendapatkan tingkat akurasi dari ringkasan sistem.

Penentuan akurasi menggunakan metode F-*measure*, *precision*, dan *recall* menurut (Baeza-Yates & Ribeiro-Neto, 1999) :

$$F_{\beta=1} = \frac{(\beta^2+1)PR}{\beta^2 P + R} = \frac{2PR}{(P+R)};\ P = \frac{|S \cap T|}{|S|};\ R = \frac{|S \cap T|}{|T|} \qquad (14)$$

S : hasil ringkasan teks dari mesin

T : hasil ringkasan teks secara manual

dengan β adalah bobot dari *precision* (P) dan *recall* (R), $\beta < 1$ penekanan pada *precision* dan $\beta > 1$ penekanan pada *recall*. Nilai *F-measure* diantara 0 dan 1, oleh karena itu nilai $\beta = 1$.

3.2 Lingkungan Pengembangan

Lingkungan yang digunakan dalam melakukan penelitian ini adalah

a. Perangkat lunak : OS Windows 10 Profesional, Compiler Active Perl versi 5.22.0.
b. Perangkat keras : AMD Richland A10-5757M 2.5 Up To 3.5 GHz, RAM 4.00 GB

3.3 Lingkungan Pengembangan

Penelitian ini dilakukan di Jurusan Ilmu Komputer Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Lampung pada semester ganjil tahun ajaran 2017-2018.

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Kesimpulan yang diperoleh dari penelitian yang telah dilakukan adalah sebagai berikut :

1. Peringkasan teks berita berbahasa Indonesia dengan menggunakan metode FFNN dan kombinasi 11 fitur teks menghasilkan rata-rata tingkat akurasi sebesar 90.9% dengan pemampatan 30%, 76.1% dengan pemampatan 20%, dan 49.1% dengan pemampatan 10%.
2. Hasil akurasi yang diperoleh pada penelitian ini lebih baik dari hasil penelitian yang dilakukan oleh (Aristoteles et al., 2012) dengan menggunakan Algoritma Genetika.
3. Hasil akurasi yang diperoleh pada penelitian (Wibowo et al., 2013) menggunakan metode Regresi Logistik Biner lebih baik dari metode FFNN.
4. Hasil analisa keterlibatan fitur teks menunjukkan bahwa fitur teks f_2, "*positive keyword*", merupakan fitur teks yang berpengaruh dan dapat merepresentasikan fitur lainnya dalam membangun ringkasan teks.
5. Jumlah fitur teks yang digunakan dalam peringkasan teks memengaruhi besar akurasi yang diperoleh.

5.2 Saran

Beberapa saran yang diperoleh setelah melakukan penelitian ini adalah sebagai berikut :

1. Mengembangkan sistem peringkasan teks dengan metode berbeda sehingga dapat menjadi perbandingan hasil yang diperoleh pada penelitian-penelitian sebelumnya dengan metode terbaru.
2. Menambahkan atau memperbarui fitur teks pada pengembangan sistem peringkasan teks untuk meningkatkan akurasi yang diperoleh sehingga dapat mencapai tingkat akurasi mendekati 100%.

DAFTAR PUSTAKA

Aristoteles, Herdiyeni, Y., Ridha, A. & Julio, A. 2012. Text Feature Weighting for Summarization of Documents in Bahasa Indonesia Using Genetic Algorithm. *International Journal of Science Issues*, pp.ISSN 1694-0814.

Baeza-Yates, R. & Ribeiro-Neto, B. 1999. *Modern Information Retrieval*. Addison-Wesley: ACM Press New York.

Budiharto, W. 2016. *Machine Learning and Computational Intelligence*. Yogyakarta: Andi.

Fattah, M.A. & Ren, F. 2008. Automatic Text Summarization. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 2.

Jolly, S.K. & Chopra, W.C.A. 2016. Text Summarization Using Neural Network Theory. *International Journal of Computer Science System*, 3(7).

Kaikhah, K. 2004. *Text Summarization Using Neural Network*. Texas: Departement of Computer Science Texas State University Texas State University.

Radev, D., Hovy, E. & McKeown, K. 2002. Introduction to The Special Issuue In Text Summarization. *Computational Linguistics*, 28, pp.399-408.

Raj, S. et al. 2015. An Artificial Neural Network Approach for Sentence Boundary Disambiguation in Urdu Language Text. *International Arab Journal of Information Technology*, 12.

Ridha, A. 2002. *Pengindeksan Otomatis dengan Istilah Tunggal untuk Dokumen Berbahasa Indonesia*. Thesis. Bogor: Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor.

Siang, J.J. 2005. *Jaringan Syaraf Tiruan dan Pemrogramannya Menggunakan MATLAB*. Yogyakarta: Andi.

Suanmali, L. Salim, N. & Binwahlan, M., 2009. Fuzzy Logic Based Method for Improving Text Summarization. *IJCSIS*, 2(1).

Wibowo, E.D., Aristoteles & Widiarti. 2013. Text Feature Weighting for Summarization of Documents Bahasa Indonesia by Using Binary Logistic Regression Algorithm. *International Joural of Computer Science and Telecommunications*, 5(7).

Yeh, J., H, K., W, Y. & I, M. 2005. Text Summarization Using A Trainable Summarizer And Latent Semantic Analysis. *Information Processing and Management*, 41(1), pp.75-95.