

**ANALISIS SENTIMEN *PUBLIC TWITTER* TERHADAP
BAKAL CALON PRESIDEN PADA PEMILU 2024
MENGUNAKAN METODE *NAIVE BAYES***

SKRIPSI



OLEH

M. VIKRY SAKIA

NIM. 1937036

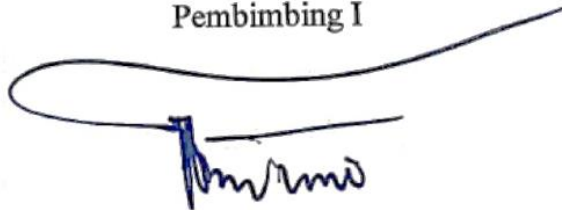
**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PASIR PENGARAIAN
2024**

PERSETUJUAN PEMBIMBING

ANALISIS SENTIMEN *PUBLIC TWITTER* TERHADAP BAKAL CALON PRESIDEN PADA PEMILU 2024 MENGGUNAKAN METODE *NAIVE* *BAYES*

Disetujui oleh :

Pembimbing I



Rivi Antoni, S.Pd., M.Pd
NIDN. 1003128103

Pembimbing II



Basorudin, S.Pd., M.Kom
NIDN. 1020088702

Deketahui Oleh :

Ketua Program Studi Teknik Informatika



Basorudin, S.Pd., M.Kom
NIDN. 1020088702

PERSETUJUAN PENGUJI

Tugas Akhir ini telah diuji oleh
Tim Penguji Ujian Sarjana Komputer
Program Studi Teknik Informatika
Fakultasi Ilmu Komputer
Universitas Pasir Pengaraian
Pada Tanggal 30 Januari 2024

Tim Penguji :

- | | |
|--|--|
| 1. <u>Rivi Antoni, S.Pd., M.Kom</u>
NIDN. 1003128103 | Ketua () |
| 2. <u>Basorudin, S.Pd., M.Kom</u>
NIDN. 1020088702 | Sekretaris () |
| 3. <u>Asep Suptiyanto, S.T., M.Kom</u>
NIDN. 1003108903 | Anggota () |
| 4. <u>Satria Riki Mustofa, S.Pd., M.Si</u>
NIDN. 1001039301 | Anggota () |
| 5. <u>Erni Rouza, S.T., M.Kom</u>
NIDN. 1009058707 | Anggota () |

Mengetahui

Dekan Fakultas Ilmu Komputer
Universitas Pasir Pengaraian



Hendri Maradona, M.Kom
NIDN. 1002038702

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa Tugas Akhir yang berjudul “Analisis Sentimen *Public Twitter* Terhadap Bakal Calon Presiden pada Pemilu 2024 Menggunakan Metode *Naive Bayes*”, benar hasil penelitian saya dengan arahan dosen pembimbing dan belum pernah diajukan dalam bentuk apapun untuk mendapatkan gelar kesarjanaan. Dalam Tugas Akhir ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasi orang lain, kecuali secara tertulis dengan jelas dicantumkan dalam naskah dengan menyebutkan referensi yang dicantumkan dalam daftar pustaka. Pernyataan ini saya buat dengan sesungguhnya dan apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena Tugas Akhir ini, serta lainnya sesuai norma yang berlaku di perguruan tinggi.

Pasir Pengaraian, 30 Januari 2024
Yang Membuat Pernyataan



M. VIKRY SAKIA
NIM : 1937036

KATA PENGANTAR

Assalamu'alaikum wa rahmatullahi wa barakatuh

Alhamdulillah Rabbil Alamin,

Segala puji syukur kehadiran Allah SWT yang senantiasa melimpahkan rahmat dan karunia-Nya, Sehingga penulis mampu menyelesaikan Laporan Magang ini dengan baik. Shalawat serta salam terucap buat junjungan kita Rasulullah Muhammad SAW karna jasa beliau yang telah membawa manusia dari zaman kebodohan ke zaman yang penuh dengan ilmu pengetahuan seperti sekarang ini.

Skripsi ini berjudul “**Analisis Sentimen *Public Twitter* Terhadap Bakal Calon Presiden Pada Pemilu 2024 Menggunakan Metode *Naive Bayes***” sebagai satu syarat untuk mendapatkan kelulusan pada jurusan Teknik Informatika Universitas Pasir Pengaraian. Banyak sekali pihak yang telah membantu penulis dalam penyusunan Skripsi ini, baik berupa bantuan materi ataupun berupa motivasi dan dukungan kepada penulis. Semua itu tentu terlalu banyak bagi penulis untuk membalasnya, namun pada kesempatan ini penulis hanya dapat mengucapkan terimakasih kepada :

1. Kepada Ayah, Ibu, dan Adikku tercinta, atas kerja keras dan selalu memberikan doa restu yang tidak ternilai harganya yang banyak memberikan semangat, motivasi, dan bimbingan yang terbaik dan limpahan kasih sayang yang tiada henti.
2. Bapak Dr. Hardianto, M.Pd selaku Rektor Universitas Pasir Pengaraian.
3. Bapak Hendri Maradona, M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas Pasir Pengaraian.
4. Bapak Basorudin, S.Pd., M.Kom selaku Ketua Prodi Teknik Informatika Fakultas Ilmu Komputer Universitas Pasir Pengaraian. Juga selaku pembimbing II yang telah memberikan bimbingan, arahan, dan saran yang berharga dalam menyusun laporan magang ini.

5. Bapak Rivi Antoni, M.Pd selaku pembimbing I yang telah memberikan bimbingan, arahan, dan saran yang berharga dalam menyusun laporan magang ini.
6. Seluruh staf dan pegawai Tata Usaha Fakultas Ilmu Komputer Universitas Pasir Pengaraian yang telah memberikan bantuan dan kelancaran administratif.
7. Kepada Nabila Ulfa, Abdus Salam dan Andika Saputra selaku teman saya yang membantu dalam pengerjaan Skripsi ini dan selalu menemani setiap saya bimbingan.
8. Kepada Pejuang Wisuda dan teman-teman seperjuangan di jurusan Teknik Informatika angkatan 2019 yang telah membantu penulis dalam menyelesaikan laporan ini didalam perkuliahan maupun diluar perkuliahan.

Penulis menyadari bahwa dalam penulisan Skripsi ini masih banyak kesalahan dan kekurangan, oleh karena itu kritik dan saran yang sifatnya membangun sangat penulis harapkan untuk kesempurnaan Skripsi ini. Akhirnya penulis berharap semoga Skripsi ini dapat memberikan sesuatu yang bermanfaat bagi siapa saja yang membacanya. Aamiin.

Wassalamu'alaikum warahmatullahi wabarakatuh

Pasir Pengaraian, 30 Januari 2024

Mahasiswa,



M. VIKRY SAKIA
NIM : 1937036

ABSTRAK

Presiden Indonesia saat ini telah memegang jabatan selama 2 periode secara berturut – turut yang mana pada dasarnya peraturan untuk menjadi calon presiden sudah tidak bisa mencalonkan kembali menjadi presiden, dalam hal itu banyak lembaga survei yang telah mengeluarkan beberapa tokoh yang memiliki elektabilitas untuk bisa menjadi calon presiden, dari hal tersebut juga banyak warganet menyampaikan pendapat, maka kesimpulan mengenai sentimen masyarakat terhadap semua tokoh bakal calon presiden tersebut menggunakan metode *naïve bayes* dan perhitungan skor sentimen dengan harapan dari penelitian ini bisa memberikan bahan referensi kepada masyarakat dalam memilih presiden di pilpres yang akan datang. Analisis sentimen adalah proses menentukan sentimen dan mengelompokkan polaritas teks dalam dokumen atau kalimat sehingga kategori dapat ditentukan sebagai sentimen positif, negatif, netral. Jejaring sosial seperti *Twitter*, umumnya digunakan dalam analisis sentimen untuk menentukan persepsi. Dari hasil pengujian akurasi yang dilakukan terhadap 750 data *training* dan 250 data *testing* diperoleh tingkat akurasi sebesar 45,6 %, dan aplikasi ini mampu mengetahui hasil komentar dari Nilai Probabilitas tertinggi dari data *testing* “Harus kah percaya tidak akan sampai terjadi permufakatan capres cawapres” yaitu sebesar 0,023409 pada $P(\text{uji}|\text{negatif})$ sehingga komentar tersebut diklasifikasikan ke dalam kelas “Negatif”. Berdasarkan hasil pengujian diketahui bahwa implementasi metode *naive bayes* melakukan pengklasifikasian data *tweet* kategori kelas sudah baik.

Kata kunci : Calon Presiden, *Naive Bayes*, Analisis Sentimen,

ABSTRACT

The current President of Indonesia has held office for 2 consecutive terms, where basically the rules for becoming a presidential candidate are no longer able to run for president again, in this case many survey institutions have issued several figures who have the electability to become a presidential candidate. , from this, many netizens also expressed their opinions, so the conclusions regarding public sentiment towards all the presidential candidate figures used the naïve Bayes method and sentiment score calculations with the hope that this research could provide reference material for the public in choosing president in the upcoming presidential election. Sentiment analysis is the process of determining sentiment and grouping the polarity of text in a document or sentence so that categories can be determined as positive, negative, neutral sentiment. Social networks such as Twitter, are commonly used in sentiment analysis to determine perceptions. From the results of accuracy testing carried out on 750 training data and 250 testing data, an accuracy level of 45.6% was obtained, and this application was able to determine the results of comments from the highest probability value from the testing data "Do you have to believe that there will be no agreement between presidential and vice presidential candidates" namely of 0.023409 in $P(\text{test}|\text{negative})$ so that the comment is classified into the "Negative" class. Based on the test results, it is known that the implementation of the Naive Bayes method for classifying tweet data in class categories is good.

Keywords : *Presidential Candidates, Naive Bayes, Sentiment Analysis*

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN PEMBIMBING	ii
LEMBAR PERSETUJUAN PENGUJI	iii
LEMBAR PERNYATAAN	iv
KATA PENGANTAR.....	v
ABSTRAK	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xiv
DAFTAR GAMBAR.....	xvi
DAFTAR SIMBOL	xix
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian.....	6
1.4 Batasan Masalah	6
1.5 Manfaat Penelitian.....	6
1.6 Sistematika Penulisan	6
BAB II LANDASAN TEORI	8
2.1 Analisis	8
2.2 Analisis Sentimen	9
2.3 <i>Public</i>	10
2.4 <i>Twiter</i>	10

2.5	Calon Presiden	11
2.6	Text Mining	11
2.7	Data Training.....	13
2.8	Preprocessing.....	13
2.9	Case Folding	13
2.10	Tokenizing	14
2.11	Stop Word	14
2.12	Normalisasi	15
2.13	Filtering	15
2.14	Stemming	15
2.15	Naïve Bayes	16
2.16	Term Frequency - Inverse Document Frequency (Tf-Idf).....	17
2.17	Aplikasi	18
2.18	UML(Unified Modeling Language).....	18
2.19	PHP (Hypertext Preprocessor).....	21
2.20	MySQL(My Structured Query Language)	21
2.21	XAMPP.....	22
2.22	Confusion Matrix.....	22
2.23	K-Fold Cross Validation	25
2.24	Flowchart	26
2.25	Penelitian Terkait	27
BAB III METODOLOGI PENELITIAN		31
3.1	Identifikasi Masalah	32
3.2	Perumusan Masalah	32
3.3	Pengumpulan Data.....	33

3.4	Analisa.....	33
3.4.1	Analisa Metode <i>Naïve Bayes</i>	33
3.4.2	Analisa Fungsi Sistem Aplikasi.....	33
3.4.3	Analisa Sistem Lama	34
3.4.4	Analisa Sistem Baru.....	34
3.5	Perancangan Sistem Aplikasi.....	34
3.5.1	Perancangan Basis Data	34
3.5.2	Perancangan Struktur Menu	34
3.5.3	Perancangan Antar Muka (<i>Interface</i>).....	34
3.6	Implementasi Sistem Aplikasi.....	35
3.7	Pengujian Aplikasi.....	35
3.7.1	Black Box Testing	35
3.8	Kesimpulan dan Saran	36
BAB IV ANALISA DAN PERANCANGAN		37
4.1	Metode Analisis	37
4.1.1	Analisa Sistem	38
4.1.2	Analisa Sistem Baru.....	38
4.1.2.1	Data Masukan (<i>Input</i>).....	39
4.1.2.2	Proses.....	40
4.1.2.3	Data Keluaran (<i>Output</i>)	40
4.2	Karakteristik Pengguna	41
4.3	Karakteristik Pengguna	41
4.3.1	Unified Modeling Language (UML)	42
4.3.1.1	Use Case Diagram	42
4.3.1.2	Class Diagram	42

4.3.1.3 Class Diagram	42
4.3.1.4 Sequence Diagram.....	47
4.4 Perancangan Subsistem Manajemen Basis Data	51
4.4.1 Rancangan Tabel <i>Login</i>	51
4.4.2 Rancangan Tabel <i>Training</i>	51
4.4.3 Rancangan Tabel <i>Testing</i>	52
4.4.4 Rancangan Tabel <i>Training Detail</i>	52
4.4.5 Rancangan Tabel <i>Testing Detail</i>	53
4.4.6 Rancangan Tabel <i>Slangword</i>	54
4.4.7 Rancangan Tabel <i>Stopword</i>	54
4.4.8 Rancangan Tabel <i>Klasifikasi</i>	54
4.4.9 Rancangan Tabel <i>Options</i>	55
4.5 Desain Sistem Secara Terinci (Perancangan Antar Muka)	55
4.5.1 Desain <i>Login</i>	55
4.5.2 Desain Halaman <i>Dashboard</i>	56
4.5.3 Desain Master Data	56
4.5.4 Desain Data <i>Training</i>	57
4.5.5 Desain <i>Preprocessing</i>	58
4.5.6 Desain Data <i>Testing</i>	59
4.5.7 Desain Akurasi	60
4.5.8 Desain Klasifikasi NBC	60
4.5.9 Desain <i>Print</i>	61
4.6 Perhitungan Manual Metode <i>Naive Bayes</i>	62
4.7 Pengujian Akurasi	72
4.8 Pengujian dengan <i>Confusion Matriks</i>	73

4.9	Pengujian Terhadap Data <i>Testing</i>	76
4.10	Pengujian <i>K-Fold Crosss Validation</i> 1-25	78
BAB V IMPLEMENTASI DAN PENGUJIAN		81
5.1	Implementasi	81
5.1.1	Batasan Implementasi	81
5.1.2	Implementasi Sistem	81
5.1.3	Hasil Implementasi	82
5.2	Pengujian Sistem	95
5.2.1	Pengujian Aplikasi Menggunakan <i>Blackbox</i>	95
5.2.1.1	Pengujian <i>Login</i>	95
5.2.1.2	Pengujian Data <i>Training</i>	96
5.2.1.3	Pengujian Data <i>Testing</i>	96
BAB VI PENUTUP		98
6.1	Kesimpulan	98
6.2	Saran	98
DAFTAR PUSTAKA		100
LAMPIRAN		

DAFTAR TABEL

Tabel 2.1 Tabel <i>Confusion Matrix</i>	24
Tabel 2.1 Skema 10 <i>Fold Cross Validation</i>	26
Tabel 2.3 Penelitian Terkait	27
Tabel 4.1 Karakteristik Pengguna	41
Tabel 4.2 <i>Login</i>	51
Tabel 4.3 Basis Data <i>Training</i>	52
Tabel 4.4 Basis Data <i>Testing</i>	52
Tabel 4.5 Basis Data <i>Training Detail</i>	53
Tabel 4.6 Basis Data <i>Testing Detail</i>	53
Tabel 4.7 Basis Data <i>Slangword</i>	54
Tabel 4.8 Basis Data <i>Stopword</i>	54
Tabel 4.9 Basis Data <i>Klasifikasi</i>	55
Tabel 4.10 Basis Data <i>Options</i>	55
Tabel 4.11 Komentar Data <i>Training</i>	62
Tabel 4.12 <i>Cleaning</i>	63
Tabel 4.13 <i>Case Folding</i>	63
Tabel 4.14 <i>Stopword</i>	63
Tabel 4.15 <i>Stemming</i>	64
Tabel 4.16 <i>Tokenizing</i>	64
Tabel 4.17 Pembobotan TF	64
Tabel 4.18 Hasil Pembobotan DF	65
Tabel 4.19 Hasil Pembobotan IDF	67
Tabel 4.20 Hasil Pembobotan TF*IDF	68

Tabel 4.21 Perhitungan Bobot Data <i>Training</i>	70
Tabel 4.22 Data Latih dan Uji	71
Tabel 4.23 Model Confusion Matriks untuk Algoritma Naïve bayes	73
Tabel 4.24 Confusion Matrix	76
Tabel 4.25 Pengujian <i>K-Fold Crosss Validation</i> 1-10	78
Tabel 4.26 Confusion Matrix Fold 1.....	79
Tabel 5.1 Pengujian <i>Login</i>.....	95
Tabel 5.2 Pengujian Data <i>Training</i>.....	96
Tabel 5.3 Pengujian Data <i>Testing</i>	96

DAFTAR GAMBAR

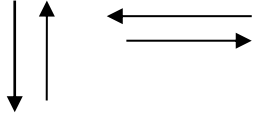
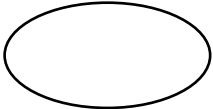
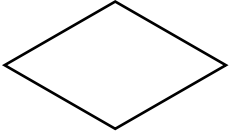
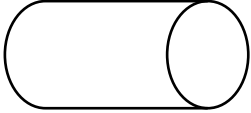
Gambar 2.1 Tahapan <i>Text Mining</i>	12
Gambar 2.2 Proses <i>Case Folding</i>	14
Gambar 2.3 Proses <i>Tokenizing</i>	14
Gambar 2.4 Proses <i>Filtering</i>	15
Gambar 3.1 Metodologi Penelitian.....	31
Gambar 4.1 <i>Waterfall</i>	37
Gambar 4.2. Analisa Sistem Baru	39
Gambar 4.3. Use Case Diagram.....	42
Gambar 4.4 <i>Class Diagram</i>	42
Gambar 4.5 Activity Diagram Login.....	43
Gambar 4.6 Activity Diagram Data Training	44
Gambar 4.7 Activity Diagram Data Testing.....	45
Gambar 4.8 Activity Diagram Preprocessing.....	46
Gambar4.9 Activity Diagram Print.....	46
Gambar 4.10 Activity Diagram Logout	47
Gambar 4.11 Sequence Diagram Login	48
Gambar 4.12 Sequence Diagram Data Training.....	48
Gambar 4.13 Sequence Diagram Data Testing	49
Gambar 4.14 Sequence Diagram Preprocessing	50
Gambar 4.15 Sequence Diagram Print	50
Gambar 4.16 Sequence Diagram Logout.....	51
Gambar 4.17 Desain <i>Login</i>	56
Gambar 4.18 Desain Halaman <i>Dashboard</i>	56

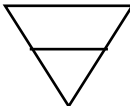
Gambar 4.19 Desain <i>Stopword</i>	57
Gambar 4.20 Desain <i>Klasifikasi</i>	57
Gambar 4.21 Desain <i>Data Training</i>	58
Gambar 4.22 Desain <i>Training</i>	58
Gambar 4.23 Desain <i>Testing</i>	59
Gambar 4.24 Desain <i>Data Testing</i>	59
Gambar 4.25 Desain <i>Akurasi</i>	60
Gambar 4.26 Desain <i>Klasifikasi NBC</i>	60
Gambar 4.27 Desain <i>Print Data Training</i>	61
Gambar 4.28 <i>Print Preprocessing Training</i>	61
Gambar 4.29 <i>Print Preprocessing Testing</i>	62
Gambar 4.30 Desain <i>Print Data Testing</i>	62
Gambar 5.1. <i>Form Login</i>	83
Gambar 5.2. <i>Form Tampilan Dashboard</i>	84
Gambar 5.3. <i>Form Stopword</i>	85
Gambar 5.4. <i>Form Klasifikasi</i>	86
Gambar 5.5. <i>Form Data Training</i>	86
Gambar 5.6. <i>Form Preprocessing Training</i>	87
Gambar 5.7. <i>Form Preprocessing Testing</i>	87
Gambar 5.8. <i>Form Data Testing</i>	88
Gambar 5.9. <i>Form Akurasi</i>	88
Gambar 5.10. <i>Form Klasifikasi NBC</i>	89
Gambar 5.11. <i>Form Hasil</i>	90
Gambar 5.12. <i>Form Input Data Training</i>	90
Gambar 5.13. <i>Form Input Data Testing</i>	91

Gambar 5.14. <i>Form Edit Data Training</i>.....	91
Gambar 5.15. <i>Form Edit Data Training</i>.....	92
Gambar 5.16. <i>Form Hasil Klasifikasi</i>.....	92
Gambar 5.17. <i>Form Hasil Print Data Training</i>.....	93
Gambar 5.18. <i>Form Hasil Print Data Testing</i>	93
Gambar 5.19. <i>Form Hasil Print Preprocessing Training</i>.....	94
Gambar 5.20. <i>Form Hasil Print Data Testing</i>	94

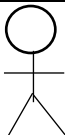
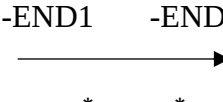
DAFTAR SIMBOL

1. Flowchart

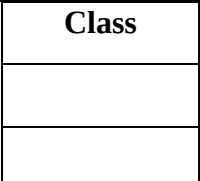
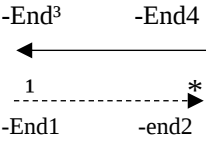
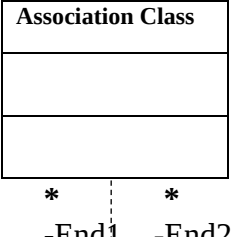
No	Simbol	Nama	Keterangan
1.		<i>Flow Direction</i>	Digunakan untuk menghubungkan antar simbol (<i>connection</i>)
2.		<i>Terminator</i>	Untuk memulai (<i>start</i>) atau akhir (<i>end</i>) dari suatu kegiatan.
3.		<i>Processing</i>	Simbol yang digunakan untuk pemrosesan suatu kegiatan.
4.		<i>Decision</i>	Pemilihan proses berdasarkan kondisi yang ada.
5.		<i>Input-Output</i>	Simbol yang menyatakan <i>input</i> dan <i>output</i> data.
6.		Dokumen	Simbol yang menyatakan <i>input</i> dan <i>output</i> yang berasal dari dokumen/hardfile berupa lembaran.
7.		<i>Database</i>	Simbol yang menyatakan <i>database</i> sistem.

	<i>Manual Operation</i>	Menggambarkan pengolahan yang tidak dilakukan komputer
	<i>Offline Storage</i>	Simbol yang menjelaskan bahwa data akan disimpan

2. *Symbol Use Case*

No	Simbol	Nama	Deskripsi
1.		<i>Case</i>	Menggambarkan proses/kegiatan yang dapat dilakukan oleh actor
2.		Aktor	Menggambarkan entitas/subyek yang dapat melakukan suatu proses
3.		<i>Relation</i>	Relasi antara case dengan aktor ataupun case dengan case lain.

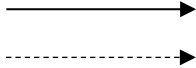
3. *Symbol Statistic Diagram*

No	Simbol	Nama	Deskripsi
1.		<i>Class</i>	Menggambarkan proses/kegiatan yang dapat dilakukan oleh aktor
2.		<i>Relation</i>	Menggambarkan hubungan komponen-komponen didalam static diagram
3.		<i>Association Class</i>	Class yang terbentuk dari hubungan antara dua buah class.

4. Simbol Sequence Diagram

No	Simbol	Nama	Deskripsi
1.		Objek	Menggambarkan pos-pos obyek yang pengirim dan penerima <i>message</i>
2.		<i>Message</i>	Menggambarkan aliran pesan yang dikirim oleh pos-pos obyek.

5. Activity Diagram

No	Simbol	Nama	Deskripsi
1.		<i>Action State</i>	Menggambarkan keadaan dari suatu elemen dalam suatu aliran aktifitas.
2.		<i>State</i>	Menggambarkan kondisi suatu elemen.
3.		<i>Control Flow</i>	Menggambarkan aliran aktifitas dari suatu elemen ke elemen lain.
4.		<i>Initial State</i>	Menggambarkan titik awal siklus hidup suatu elemen.
5.		<i>Final State</i>	Menggambarkan titik akhir yang menjadi kondisi akhir suatu elemen.