BAB 3
ANALISIS DAN PERANCANGAN SISTEM

3.1 Analisis Masalah

Penyebaran dan kebutuhan akan informasi digital di Indonesia dalam bentuk teks atau dokumen semakin meningkat dan setiap waktu terus mengalami pertumbuhan seiring dengan perkembangan teknologi. Salah satu sumber informasi tersebut adalah portal berita elektronik. Suatu portal berita elektronik mengklasifikasi artikel – artikel ke dalam kategori. Selama ini pengklasifikasian berita masih menggunakan tenaga manusia atau manual. Kategori yang banyak beserta waktu yang cepat akan menyulitkan editor untuk mengklasifikasikan artikel, terutama pada artikel yang isinya tidak terlalu berbeda secara jelas. Beberapa kategori yang penggunaan bahasanya tidak berbeda terlalu jauh seperti olahraga, sains, ekonomi, teknologi dan kesehatan mengharuskan seorang editor mengetahui isi artikel secara keseluruhan untuk selanjutnya dimasukkan ke dalam kategori yang tepat. Akan lebih efesien apabila kategori berita dimasukkan secara otomatis dengan menggunakan metode tertentu. Pengklasifikasian juga dilakukan untuk mempermudah para pengguna dalam mengakses artikel.

Support Vector Machine (SVM) adalah sistem pembelajaran yang menggunakan ruang hipotesis berupa fungsi-fungsi linier dalam sebuah ruang fitur (*feature space*) berdimensi tinggi, dilatih dengan algoritma pembelajaran yang didasarkan pada teori optimasi dengan mengimplementasikan learning bias yang berasal dari teori pembelajaran statistik [1]. SVM memiliki kelebihan yaitu mampu menemukan fungsi pemisah (klasifier) yang optimal yang bisa memisahkan dua set data dari dua kelas yang berbeda. Tetapi SVM pada saat pertama kali diperkenalkan oleh Vapnik hanya dapat mengklasifikasikan data ke dalam dua kelas (klasifikasi biner). Sementara masalah di dunia nyata umumnya mempunyai banyak kelas. Salah satu cara untuk mengimplementasikan *Multiclass* SVM yaitu dengan menggabungkan beberapa SVM biner [1]. Pada pengaturan *Multiclass*, sebuah dokumen teks masuk dalam tepat salah satu kelas dari beberapa kelas. Pendekatan *Multiclass* SVM yang digunakan dalam penelitian adalah *One-vs-One*, *One-vs-All*,

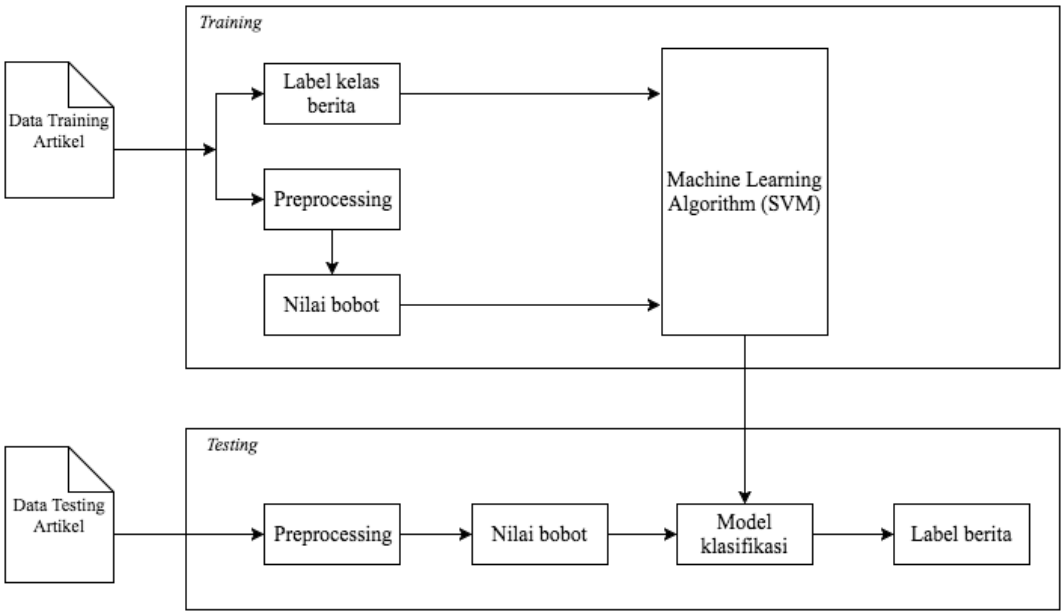
dan *error correcting output code* (ECOC).

Pada penelitian sebelumnya sudah ada yang melakukan perbandingan pendekatan *Multiclass* SVM. Pendekatan yang digunakan adalah *one vs one*, *one vs all* [20]. Data yang digunakan dalam penelitian sebelumnya adalah data aroma yang terdiri dari 3 jenis aroma. Dari hasil percobaan memberikan hasil bahwa metode *One-vs-One* telah mampu 100% mengklasifikasikan data aroma berdasarkan kelas yang tepat. Semakin banyak data training yang digunakan, metode *One-vs-One* akan lebih cepat mengklasifikasikan data dibandingkan dengan metode *One-vs-Rest* [20].

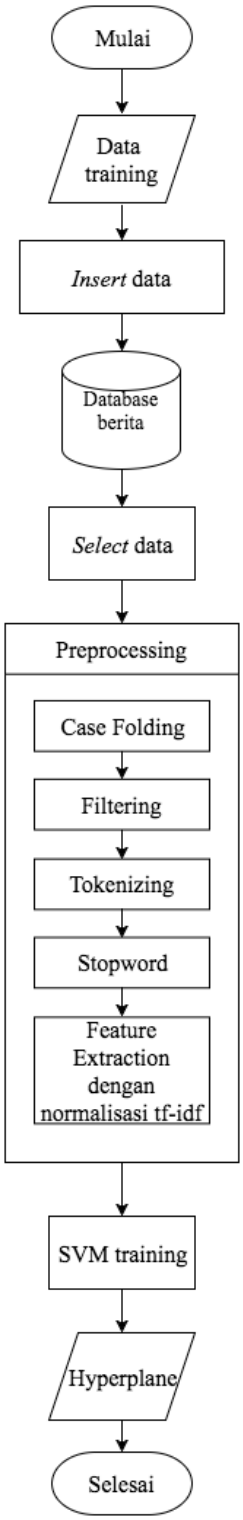
Selain metode *One vs One* dan *One vs All*, ada metode *error correcting output code* (ECOC). Pendekatan *error correcting output code* (ECOC) merupakan pendekatan yang terinspirasi dari pendekatan teori informasi untuk mengirimkan pesan melalui saluran yang ber-noise [1]. Metode *error correcting output code* (ECOC) dianggap sebagai varian dari metode *One vs All classification* [21]. Tetapi belum ada penelitian yang membandingkan performansi pada metode *One vs One* dan *One vs All* dengan metode ECOC khususnya pada data berbahasa Indonesia, padahal ECOC merupakan salah satu metode yang paling popular [10]. Oleh karena itu dilakukan perbandingan metode *One vs All, One vs One* dan ECOC untuk mengklasifikasikan dokumen artikel.

3.1.1 Analisis Sistem

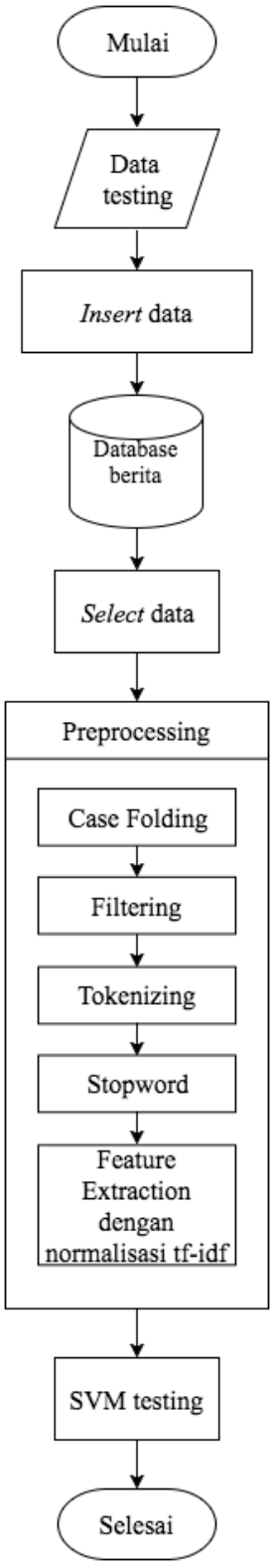
Klasifikasi artikel dengan membandingkan metode *One vs All, One vs One* dan ECOC pada *Multiclass* SVM menjadi permasalahan yang akan dibahas pada penelitian ini. Gambar 3.1 menggambarkan arsitektur sistem secara umum. Dalam membangun sistem klasifikasi artikel, dilakukan beberapa tahapan analisis. Sistem ini menerima masukan berupa data teks dengan file format .csv. Data teks ini akan akan dilakukan *preprocessing* terlebih dahulu sebelum dilakukan tahap pembelajaran dan tahap klasifikasi. Berikut tahapan yang dilakukan yang dapat dilihat pada Gambar 3.2 dan 3.3.



Gambar 3.1 Arsitektur Sistem



Gambar 3.2 *Flowchart* Sistem *Training*



Gambar 3.3 *Flowchart* Sistem *Testing*

3.1.2 Analisis Data Masukan

Dalam melakukan percobaan klasifikasi teks ini akan digunakan data masukan merupakan data artikel yang berasal dari internet. Data tersebut diambil secara random sesuai kategori yang diambil dari:

1. www.tempo.com
2. www.kompas.com
3. www.merdeka.com
4. www.tribunnews.com

Jumlah data yang diambil dari url diatas kurang lebih 1500 data training dan 200 data uji. Semua data yang diambil merupakan data berita tahun 2016. Format data yang digunakan sebagai input adalah .csv. Dengan isi berupa kumpulan kalimat atau berupa uraian karangan yang sesuai dengan fakta. Didalam input tidak disertai dengan gambar ataupun karakter HTML dan tidak ada typo pada teks. Berikut merupakan contoh artikel yang akan digunakan sebagai data training, data yang digunakan sebanyak 5 data.

Artikel	Kelas
Saham Boeing Co bisa jatuh hingga 15 persen jika penjualan pesawat melambat akibat lemahnya permintaan. Hal ini dipaparkan dalam publikasi laporan keuangan Barron pada Minggu (22/5/2016). Publikasi tersebut juga mengatakan, maskapai penerbangan dan perusahaan leasing lebih memilih menggunakan pesawat lama dan menunda order untuk pesawat baru.	ekonomi
Demensia adalah kumpulan gangguan dengan gejala yang mengakibatkan perubahan pada pasien dalam cara mereka berpikir dan berinteraksi dengan orang lain. Seringkali, memori jangka pendek, pikiran dan kemampuan bicara dan motorik penderitanya terpengaruh. Demensia kebanyakan diderita oleh orang-orang lanjut usia. Merawat dan mendampingi orang dengan demensia, kadang terasa sulit jika Anda tidak tahu apa yang harus Anda lakukan.	kesehatan
Sejumlah pengguna layanan pesan instan WhatsApp di platform iOS (iPhone) mengeluhkan tidak bisa mengakses layanan tersebut pada hari ini, Kamis (26/5/2016). Pantauan, keluhan itu sudah muncul sejak Rabu (25/5/2016) sore kemarin. Sementara pengguna WhatsApp di platform nampaknya tidak menemui kendala yang sama.	teknologi

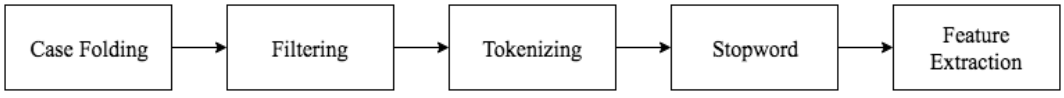
Pemerintah Kabupaten Banyuwangi menggelar Agro Expo 2016 dan memamerkan produk-produk pertanian asli Banyuwangi mulai dari tanaman pangan sampai holtikultura di Taman Blambangan. Event ini adalah bentuk perhatian pemerintah kepada petani dan menjadi panggung dari para petani untuk menampilkan produk unggulannya, kata Bupati Banyuwangi Abdullah Azwar Anas kepada KompasTravel, Senin (11/4/2016). Menurut Anas, selama lima tahun terakhir Pemerintah Kabupaten Banyuwangi telah menetapkan kebijakan proteksi buah lokal dengan mewajibkan suguhan buah lokal untuk acara yang melibatkan pemerintah kabupaten termasuk PNS yang akan memberi oleh-oleh buah.	Pariwisata
Ganda puteri Greysia Polii/Nitya Krishinda Maheswari akhirnya sukses menumbangkan pasangan Tiongkok, Yu Yang/Tang Yuanting. Mereka menang dua gim langsung di perempat final India terbuka Super Series 2016, 21-19 dan 21-12.Pasangan Indonesia ini mampu memimpin jalannya pertandingan sejak menit pertama. Dari gim pertama hingga game kedua, mereka berhasil mementahkan serangan lawan dan ungguli perolehan poin.Ini merupakan kemenangan pertama Greysia/Nitya atas Yu/Tang.	olahraga

3.1.3 Analisis Proses

Proses pendekatan sistematis untuk mengidentifikasi permasalahan, peluang dan tujuan dalam merancang suatu sistem.

3.1.3.1 *Preprocessing*

Tujuan dari tahap ini adalah menghilangkan noise, menyeragamkan bentuk kata dan mengurangi volume kosa kata. Tahap ini meliputi *case folding*, *filtering*, *tokenizing*, *stopwords* dan *feature extraction*. Proses *preprocessing* ini akan mempengaruhi hasil klasifikasi karena dari hasil *preprocessing* ini akan dilakukan proses pelatihan yang kemudian akan dilanjutkan dengan proses pengujian klasifikasi. Berikut gambaran tahapan praproses dapat dilihat pada gambar 3.2.



Gambar 3.4 Flowchart *preprocessing*

3.1.3.2 *Case Folding*

Pada tahap ini dilakukan pengubahan huruf di dalam dokumen menjadi huruf kecil. Proses *case folding* dapat dilihat pada gambar 3.5 :



Gambar 3.5 *Flowchart* tahapan *case folding*

Berikut ini adalah contoh dokumen yang telah melewati tahapan *case folding* pada Tabel 3.1.

Tabel 3.1 Contoh tahapan *Case folding*

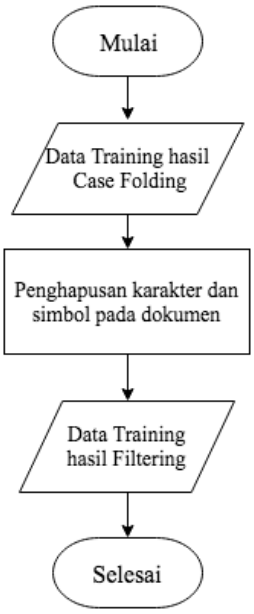
Input	Output
Saham Boeing Co bisa jatuh hingga 15 persen jika penjualan pesawat melambat akibat lemahnya permintaan. Hal ini dipaparkan dalam publikasi laporan keuangan Barron pada Minggu (22/5/2016). Publikasi tersebut juga mengatakan, maskapai penerbangan dan perusahaan leasing lebih memilih menggunakan pesawat lama dan menunda order untuk pesawat baru.	saham boeing co bisa jatuh hingga 15 persen jika penjualan pesawat melambat akibat lemahnya permintaan. hal ini dipaparkan dalam publikasi laporan keuangan barron pada minggu (22/5/2016). publikasi tersebut juga mengatakan, maskapai penerbangan dan perusahaan leasing lebih memilih menggunakan pesawat lama dan menunda order untuk pesawat baru.
Demensia adalah kumpulan gangguan dengan gejala yang mengakibatkan	demensia adalah kumpulan gangguan dengan gejala yang mengakibatkan

perubahan pada pasien dalam cara mereka berpikir dan berinteraksi dengan orang lain. Seringkali, memori jangka pendek, pikiran dan kemampuan bicara dan motorik penderitanya terpengaruh. Demensia kebanyakan diderita oleh orang-orang lanjut usia. Merawat dan mendampingi orang dengan demensia, kadang terasa sulit jika Anda tidak tahu apa yang harus Anda lakukan.	perubahan pada pasien dalam cara mereka berpikir dan berinteraksi dengan orang lain. seringkali, memori jangka pendek, pikiran dan kemampuan bicara dan motorik penderitanya terpengaruh. demensia kebanyakan diderita oleh orang-orang lanjut usia. merawat dan mendampingi orang dengan demensia, kadang terasa sulit jika anda tidak tahu apa yang harus anda lakukan.
Sejumlah pengguna layanan pesan instan WhatsApp di platform iOS (iPhone) mengeluhkan tidak bisa mengakses layanan tersebut pada hari ini, Kamis (26/5/2016). Pantauan, keluhan itu sudah muncul sejak Rabu (25/5/2016) sore kemarin. Sementara pengguna WhatsApp di platform nampaknya tidak menemui kendala yang sama.	sejumlah pengguna layanan pesan instan whatsapp di platform ios (iphone) mengeluhkan tidak bisa mengakses layanan tersebut pada hari ini, kamis (26/5/2016). pantauan, keluhan itu sudah muncul sejak rabu (25/5/2016) sore kemarin. sementara pengguna whatsapp di platform nampaknya tidak menemui kendala yang sama.
Pemerintah Kabupaten Banyuwangi menggelar Agro Expo 2016 dan memamerkan produk-produk pertanian asli Banyuwangi mulai dari tanaman pangan sampai holtikultura di Taman Blambangan. Event ini adalah bentuk perhatian pemerintah kepada petani dan menjadi panggung dari para petani untuk menampilkan produk unggulannya, kata Bupati Banyuwangi Abdullah Azwar Anas kepada KompasTravel, Senin (11/4/2016). Menurut Anas, selama lima tahun terakhir Pemerintah Kabupaten Banyuwangi telah menetapkan kebijakan proteksi buah lokal dengan mewajibkan suguhan buah lokal untuk acara yang melibatkan pemerintah kabupaten termasuk PNS yang akan memberi oleh-oleh buah.	pemerintah kabupaten banyuwangi menggelar agro expo 2016 dan memamerkan produk-produk pertanian asli banyuwangi mulai dari tanaman pangan sampai holtikultura di taman blambangan. event ini adalah bentuk perhatian pemerintah kepada petani dan menjadi panggung dari para petani untuk menampilkan produk unggulannya, kata bupati banyuwangi abdullah azwar anas kepada kompas travel, senin (11/4/2016). menurut anas, selama lima tahun terakhir pemerintah kabupaten banyuwangi telah menetapkan kebijakan proteksi buah lokal dengan mewajibkan suguhan buah lokal untuk acara yang melibatkan pemerintah kabupaten termasuk pns yang akan memberi oleh-oleh buah.
Ganda puteri Greysia Polii/Nitya Krishinda Maheswari akhirnya sukses menumbangkan pasangan Tiongkok, Yu Yang/Tang Yuanting. Mereka menang dua	ganda puteri greysia polii/nitya krishinda maheswari akhirnya sukses menumbangkan pasangan tiongkok, yu yang/tang yuanting. mereka menang dua

gim langsung di perempat final India terbuka Super Series 2016, 21-19 dan 21-12.Pasangan Indonesia ini mampu memimpin jalannya pertandingan sejak menit pertama. Dari gim pertama hingga game kedua, mereka berhasil mementahkan serangan lawan dan ungguli perolehan poin.Ini merupakan kemenangan pertama Greysia/Nitya atas Yu/Tang.	gim langsung di perempat final india terbuka super series 2016, 21-19 dan 21-12. pasangan indonesia ini mampu memimpin jalannya pertandingan sejak menit pertama. dari gim pertama hingga game kedua, mereka berhasil mementahkan serangan lawan dan ungguli perolehan poin. ini merupakan kemenangan pertama greysia/nitya atas yu/tang.

3.1.3.3 *Filtering*

Filtering digunakan untuk menghilangkan karakter – karakter non – huruf yang dianggap tidak valid yaitu karakter lain selain karakter "a" sampai "z" dan spasi. Proses filtering ini mempunyai alur yang dapat dilihat pada Gambar 3.6. Sedangkan, karakter – karakter yang akan dihapus atau dianggap sebagai pemisah kata, dapat dilihat pada Tabel 3.2.



Gambar 3.6 Flowchart *Filtering*

Tabel 3.2 merupakan contoh karakter yang akan dihapus dalam proses *filtering.*

Tabel 3.2 Karakter yang akan dihapus

Karakter			
!	(	<	]
@	)	:	|
#	.	;	\
$	,	/	+
%	‘	?	=
^	“	{	_
&	-	}	,
*	>	[	Tab, spasi, enter

Berikut ini adalah contoh dokumen yang telah melewati tahapan *filtering* yang akan digambarkan pada Tabel 3.3.

Tabel 3.3 Contoh tahapan *Filtering*

Data hasil *case folding*	Output
saham boeing co bisa jatuh hingga 15 persen jika penjualan pesawat melambat akibat lemahnya permintaan, hal ini dipaparkan dalam publikasi laporan keuangan barron pada minggu (22/5/2016). publikasi tersebut juga mengatakan, maskapai penerbangan dan perusahaan leasing lebih memilih menggunakan pesawat lama dan menunda order untuk pesawat baru.	saham boeing co bisa jatuh hingga persen jika penjualan pesawat melambat akibat lemahnya permintaan hal ini dipaparkan dalam publikasi laporan keuangan barron pada minggu publikasi tersebut juga mengatakan maskapai penerbangan dan perusahaan leasing lebih memilih menggunakan pesawat lama dan menunda order untuk pesawat baru
demensia adalah kumpulan gangguan dengan gejala yang mengakibatkan perubahan pada pasien dalam cara mereka berpikir dan berinteraksi dengan orang lain, seringkali, memori jangka pendek, pikiran dan kemampuan bicara dan motorik penderitanya terpengaruh, demensia kebanyakan diderita oleh orang - orang lanjut usia, merawat dan mendampingi orang dengan demensia, kadang terasa	demensia adalah kumpulan gangguan dengan gejala yang mengakibatkan perubahan pada pasien dalam cara mereka berpikir dan berinteraksi dengan orang lain seringkali memori jangka pendek pikiran dan kemampuan bicara dan motorik penderitanya terpengaruh. demensia kebanyakan diderita oleh orang orang lanjut usia merawat dan mendampingi orang dengan demensia kadang terasa sulit

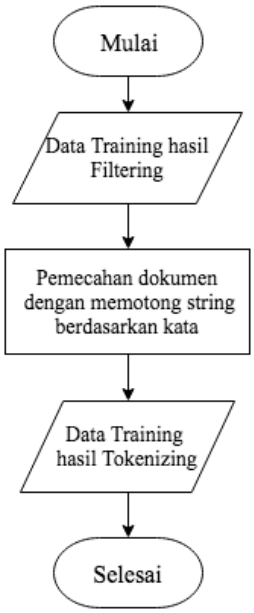
sulit jika anda tidak tahu apa yang harus anda lakukan.	jika anda tidak tahu apa yang harus anda lakukan
sejumlah pengguna layanan pesan instan whatsapp di platform ios (iPhone) mengeluhkan tidak bisa mengakses layanan tersebut pada hari ini, kamis (26/5/2016). pantauan, keluhan itu sudah muncul sejak rabu (25/5/2016) sore kemarin. sementara pengguna whatsapp di platform nampaknya tidak menemui kendala yang sama.	sejumlah pengguna layanan pesan instan whatsApp di platform ios iphone mengeluhkan tidak bisa mengakses layanan tersebut pada hari ini kamis pantauan keluhan itu sudah muncul sejak rabu sore kemarin sementara pengguna whatsapp di platform nampaknya tidak menemui kendala yang sama
pemerintah kabupaten banyuwangi menggelar agro expo 2016 dan memamerkan produk - produk pertanian asli banyuwangi mulai dari tanaman pangan sampai holtikultura di taman blambangan. event ini adalah bentuk perhatian pemerintah kepada petani dan menjadi panggung dari para petani untuk menampilkan produk unggulannya, kata bupati banyuwangi abdullah azwar anas kepada kompas travel, senin (11/4/2016). menurut anas, selama lima tahun terakhir pemerintah kabupaten banyuwangi telah menetapkan kebijakan proteksi buah lokal dengan mewajibkan suguhan buah lokal untuk acara yang melibatkan pemerintah kabupaten termasuk pns yang akan memberi oleh-oleh buah.	pemerintah kabupaten banyuwangi menggelar agro expo dan memamerkan produk produk pertanian asli banyuwangi mulai dari tanaman pangan sampai holtikultura di taman blambangan event ini adalah bentuk perhatian pemerintah kepada petani dan menjadi panggung dari para petani untuk menampilkan produk unggulannya, kata bupati banyuwangi abdullah azwar anas kepada kompas travel senin menurut anas selama lima tahun terakhir pemerintah kabupaten banyuwangi telah menetapkan kebijakan proteksi buah lokal dengan mewajibkan suguhan buah lokal untuk acara yang melibatkan pemerintah kabupaten termasuk pns yang akan memberi oleh-oleh buah
ganda puteri greysia polii/nitya krishinda maheswari akhirnya sukses menumbangkan pasangan tiongkok, yu yang/tang yuanting. mereka menang dua gim langsung di perempat final india terbuka super series 2016, 21-19 dan 21-12. pasangan indonesia ini mampu memimpin jalannya pertandingan sejak menit pertama. dari gim pertama hingga game kedua, mereka berhasil mementahkan serangan lawan dan ungguli	ganda puteri greysia polii nitya krishinda maheswari akhirnya sukses menumbangkan pasangan tiongkok yu yang tang yuanting mereka menang dua gim langsung di perempat final india terbuka super series dan pasangan indonesia ini mampu memimpin jalannya pertandingan sejak menit pertama. dari gim pertama hingga game kedua mereka berhasil mementahkan serangan lawan dan ungguli perolehan poin ini merupakan

perolehan poin. ini merupakan kemenangan pertama greysia/nitya atas yu/tang.	kemenangan pertama greysia nitya atas yu tang

3.1.3.4 *Tokenizing*

Tahap *tokenizing/parsing* merupakan proses pemotongan string input berdasarkan tiap kata yang menyusunya serta membedakan karakter-karakter tertentu yang dapat diperlakukan sebagai pemisah kata atau bukan.

Proses tokenizing ini mempunyai alur yang digambarkan pada gambar 3.7 sebagai berikut.



Gambar 3.7 *Flowchart* tahapan *Tokenizing*

Berikut ini adalah contoh dokumen yang telah melewati tahapan *tokenizing* yang akan di ilustrasikan pada Tabel 3.4.

Tabel 3.4 Contoh tahapan *Tokenizing*

Data hasil *Filtering*	Output
saham boeing co bisa jatuh hingga persen jika penjualan pesawat melambat akibat lemahnya permintaan hal ini dipaparkan dalam publikasi laporan keuangan barron pada minggu publikasi	'saham', 'boeing', 'co', 'bisa', 'jatuh', 'hingga', 'persen', 'jika', 'penjualan', 'pesawat', 'melambat', 'akibat', 'lemahnya', 'permintaan', 'hal', 'ini', 'dipaparkan', 'dalam', 'publikasi',

tersebut juga mengatakan maskapai penerbangan dan perusahaan leasing lebih memilih menggunakan pesawat lama dan menunda order untuk pesawat baru	'laporan', 'keuangan', 'barron', 'pada', 'minggu', 'publikasi', 'tersebut', 'juga', 'mengatakan', 'maskapai', 'penerbangan', 'dan', 'perusahaan', 'leasing', 'lebih', 'memilih', 'menggunakan', 'pesawat', 'lama', 'dan', 'menunda', 'order', 'untuk', 'pesawat', 'baru'
demensia adalah kumpulan gangguan dengan gejala yang mengakibatkan perubahan pada pasien dalam cara mereka berpikir dan berinteraksi dengan orang lain seringkali memori jangka pendek pikiran dan kemampuan bicara dan motorik penderitanya terpengaruh. demensia kebanyakan diderita oleh orang-orang lanjut usia merawat dan mendampingi orang dengan demensia, kadang terasa sulit jika anda tidak tahu apa yang harus anda lakukan	'pada', 'sulit', 'gejala', 'penderitanya', 'gangguan', 'anda', 'lakukan', 'orang', 'seringkali', 'perubahan', 'apa', 'bicara', 'usia', 'cara', 'memori', 'dalam', 'mereka', 'harus', 'dengan', 'adalah', 'lain', 'kumpulan', 'kemampuan', 'terasa', 'mengakibatkan', 'pikiran', 'terpengaruh', 'dan', 'mendampingi', 'pasien', 'jangka', 'berinteraksi', 'lanjut', 'kadang', 'kebanyakan', 'motorik', 'tidak', 'demensia', 'tahu', 'pendek', 'diderita', 'jika', 'merawat', 'yang', 'berpikir', 'oleh'
sejumlah pengguna layanan pesan instan whatsApp di platform ios iphone mengeluhkan tidak bisa mengakses layanan tersebut pada hari ini kamis pantauan keluhan itu sudah muncul sejak rabu sore kemarin sementara pengguna whatsapp di platform nampaknya tidak menemui kendala yang sama	'pada', 'pesan', 'muncul', 'hari', 'kendala', 'pengguna', 'mengeluhkan', 'ini', 'kemarin', 'itu', 'kamis', 'sejak', 'instan', 'rabu', 'sama', 'nampaknya', 'sudah', 'platform', 'sementara', 'iphone', 'whatsapp', 'menemui', 'di', 'ios', 'sore', 'tersebut', 'mengakses', 'tidak', 'keluhan', 'sejumlah', 'yang', 'pantauan', 'layanan', 'bisa'
pemerintah kabupaten banyuwangi menggelar agro expo dan memamerkan produk produk pertanian asli banyuwangi mulai dari tanaman pangan sampai holtikultura di taman blambangan event ini adalah bentuk perhatian pemerintah kepada petani dan menjadi panggung dari	'pemerintah', 'kabupaten', 'mulai', 'memamerkan','menggelar','menetapkan', 'menurut', 'agro', 'banyuwangi', 'terakhir', 'ini', 'suguhan', 'acara', 'kata', 'lima', 'event', 'petani', 'bupati', 'termasuk', 'pangan', 'unggulannya', 'taman', 'abdullah', 'menampilkan', 'azwar', 'mewajibkan', 'expo', 'holtikultura',

para petani untuk menampilkan produk unggulannya, kata bupati banyuwangi abdullah azwar anas kepada kompas travel senin menurut anas selama lima tahun terakhir pemerintah kabupaten banyuwangi telah menetapkan kebijakan proteksi buah lokal dengan mewajibkan suguhan buah lokal untuk acara yang melibatkan pemerintah kabupaten termasuk pns yang akan memberi oleh-oleh buah	'sampai', 'adalah', 'panggung', 'akan', 'dengan', 'para', 'di', 'dan', 'memberi', 'kebijakan', 'dari', 'tahun', 'asli', 'kepada', 'pertanian', 'bentuk', 'perhatian', 'blambangan', 'senin', 'anas', 'tanaman', 'buah', 'selama', 'pns', 'melibatkan', 'menjadi', 'untuk', 'telah', 'yang', 'lokal', 'produk', 'kompastravel', 'oleh', 'proteksi'
ganda puteri greysia polii nitya krishinda maheswari akhirnya sukses menumbangkan pasangan tiongkok yu yang tang yuanting mereka menang dua gim langsung di perempat final india terbuka super series dan pasangan indonesia ini mampu memimpin jalannya pertandingan sejak menit pertama. dari gim pertama hingga game kedua mereka berhasil mementahkan serangan lawan dan ungguli perolehan poin ini merupakan kemenangan pertama greysia nitya atas yu tang	'indonesia', 'series', 'memimpin', 'india', 'lawan', 'kemenangan', 'dari', 'menumbangkan', 'pertandingan', 'menit', 'kedua', 'puteri', 'mereka', 'yangtang', 'sukses', 'dua', 'akhirnya', 'yutang', 'langsung', 'menang', 'mampu', 'gim', 'sejak', 'mementahkan', 'perolehan', 'perempat', 'hingga', 'di', 'dan', 'berhasil', 'game', 'serangan', 'krishinda', 'ungguli', 'atas', 'super', 'pasangan', 'yu', 'terbuka', 'merupakan', 'jalannya', 'pertama', 'yuanting', 'poliinitya', 'ganda', 'final', 'tiongkok', 'greysia', 'poinini', 'greysianitya', 'maheswari', 'ini'

3.1.3.5 *Stopword*

Tahap *stopword* adalah tahap mengambil kata – kata yang tidak begitu memiliki makna pada teks dari hasil *tokenizing*. Biasanya kata-kata yang sering muncul atau kata-kata tersebut tidak memiliki arti atau tidak relevan, Seperti kata sambung, kata depan, kata ganti, kata penghubung, dan sebagainya. Contohnya pada kata “selalu” akan dibandingkan dengan kata yang sudah ada pada kamus *stopword* yang telah dibuat. Karena kata tersebut ada dalam kamus *stopword* maka kata tersebut akan dihapus.



Gambar 3.8 *Flowchart Stopwords*

Pada tahapan ini, ada beberapa aturan proses agar hasil *stopword* sesuai apa yang diharapkan. Untuk daftar *stopword* yang digunakan berdasarkan usulan dari tesis Talla. *Stopword* yang diusulkan berasal dari hasil analisis frekuensi kata dalam Bahasa Indonesia [15]. Hal ini dibandingkan dengan hasil penelitian linguistik komputasi dalam Bahasa Indonesia [18] dan dengan stoplist di [19]. Adapun isi dari sebagian daftar kata *stopword* disajikan pada Tabel 3.5, sedangkan untuk daftar lengkap *stopword* dapat dilihat pada lampiran C.

Tabel 3.5 Daftar Sebagian *Stopword*

telah	punya	mendapatkan	dari
antara	agar	juga	semua
setelah	jadi	ada	setelah

Berikut ini adalah contoh dokumen yang telah melewati tahapan *tokenizing* yang akan di ilustrasikan pada Tabel 3.6.

Tabel 3.6 Contoh tahapan *Stopword*

Data hasil *tokenizing*	Hasil *Stopword*
'saham', 'boeing', 'co', 'bisa', 'jatuh', 'hingga', 'persen', 'jika', 'penjualan', 'pesawat', 'melambat', 'akibat', 'lemahnya', 'permintaan', 'hal', 'ini', 'dipaparkan', 'dalam', 'publikasi', 'laporan', 'keuangan', 'barron', 'pada', 'minggu', 'publikasi', 'tersebut', 'juga', 'mengatakan', 'maskapai', 'penerbangan', 'dan', 'perusahaan', 'leasing', 'lebih', 'memilih', 'menggunakan', 'pesawat', 'lama', 'dan', 'menunda', 'order', 'untuk', 'pesawat', 'baru'	'persen', 'leasing', 'penjualan', 'barron', 'jatuh', 'pesawat', 'keuangan', 'saham', 'boeing', 'laporan', 'maskapai', 'perusahaan', 'co', 'lemahnya', 'penerbangan', 'minggu', 'akibat','memilih', 'publikasi', 'dipaparkan', 'permintaan', 'melambat', 'menunda', 'order', ‘pesawat’
'pada', 'sulit', 'gejala', 'penderitanya', 'gangguan', 'anda', 'lakukan', 'orang', 'seringkali', 'perubahan', 'apa', 'bicara', 'usia', 'cara', 'memori', 'dalam', 'mereka', 'harus', 'dengan', 'adalah', 'lain', 'kumpulan', 'kemampuan', 'terasa', 'mengakibatkan', 'pikiran', 'terpengaruh', 'dan', 'mendampingi', 'pasien', 'jangka', 'berinteraksi', 'lanjut', 'kadang', 'kebanyakan', 'motorik', 'tidak', 'demensia', 'tahu', 'pendek', 'diderita', 'jika', 'merawat', 'yang', 'berpikir', 'oleh'	'sulit', 'gejala', 'penderitanya', 'pikiran', 'gangguan', 'mendampingi', 'lakukan', 'orang', 'seringkali', 'pasien', 'perubahan', 'jangka', 'berinteraksi', 'mengakibatkan', 'kadang', 'kebanyakan', 'motorik', 'bicara', 'usia', 'demensia', 'memori', 'pendek', 'diderita', 'merawat', 'terpengaruh', 'berpikir', 'kumpulan', 'kemampuan'
'pada', 'pesan', 'muncul', 'hari', 'kendala', 'pengguna', 'mengeluhkan', 'ini', 'kemarin', 'itu', 'kamis', 'sejak', 'instan', 'rabu', 'sama', 'nampaknya', 'sudah', 'platform', 'sementara', 'iphone', 'whatsapp', 'menemui', 'di', 'ios', 'sore', 'tersebut', 'mengakses', 'tidak', 'keluhan', 'sejumlah', 'yang', 'pantauan', 'layanan', 'bisa'	'pesan', 'muncul', 'ios', 'kendala', 'pengguna', 'mengeluhkan', 'sore', 'kemarin', 'mengakses', 'instan', 'kamis', 'menemui', 'rabu', 'keluhan', 'nampaknya', 'platform', 'pantauan', 'layanan', 'whatsapp', 'iphone'

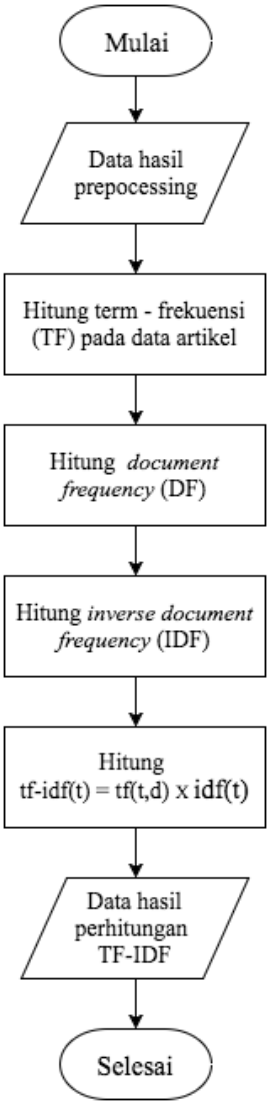
'pemerintah', 'kabupaten', 'mulai', 'memamerkan','menggelar','menetapkan', 'menurut', 'agro', 'banyuwangi', 'terakhir', 'ini', 'suguhan', 'acara', 'kata', 'lima', 'event', 'petani', 'bupati', 'termasuk', 'pangan', 'unggulannya', 'taman', 'abdullah', 'menampilkan', 'azwar', 'mewajibkan', 'expo', 'holtikultura', 'sampai', 'adalah', 'panggung', 'akan', 'dengan', 'para', 'di', 'dan', 'memberi', 'kebijakan', 'dari', 'tahun', 'asli', 'kepada', 'pertanian', 'bentuk', 'perhatian', 'blambangan', 'senin', 'anas', 'tanaman', 'buah', 'selama', 'pns', 'melibatkan', 'menjadi', 'untuk', 'telah', 'yang', 'lokal', 'produk', 'kompastravel', 'oleh', 'proteksi'	'pemerintah','kabupaten','memamerkan', 'menggelar', 'pertanian', 'menetapkan', 'agro', 'banyuwangi', 'panggung', 'suguhan', 'acara', 'event', 'bupati', 'pangan', 'unggulannya', 'taman', 'abdullah', 'menampilkan', 'azwar', 'mewajibkan', 'expo', 'holtikultura', 'petani', 'pns', 'kebijakan', 'perhatian', 'bentuk', 'blambangan', 'senin', 'tanaman', 'buah', 'anas', 'melibatkan', 'asli', 'lokal', 'produk', 'kompastravel', 'proteksi'
'indonesia', 'series', 'memimpin', 'india', 'lawan', 'kemenangan', 'dari', 'menumbangkan', 'pertandingan', 'menit', 'kedua', 'puteri', 'mereka', 'yangtang', 'sukses', 'dua', 'akhirnya', 'yutang', 'langsung', 'menang', 'mampu', 'gim', 'sejak', 'mementahkan', 'perolehan', 'perempat', 'hingga', 'di', 'dan', 'berhasil', 'game', 'serangan', 'krishinda', 'ungguli', 'atas', 'super', 'pasangan', 'yu', 'terbuka', 'merupakan', 'jalannya', 'pertama', 'yuanting', 'poliinitya', 'ganda', 'final', 'tiongkok', 'greysia', 'poinini', 'greysianitya', 'maheswari', 'ini'	'series', 'memimpin', 'india', 'yuanting', 'lawan','kemenangan', 'menumbangkan', 'pertandingan', 'menit', 'puteri', 'yangtang', 'sukses', 'super', 'yutang', 'langsung', 'menang', 'gim', 'mementahkan', 'perolehan', 'perempat', 'berhasil', 'ganda', 'serangan', 'krishinda', 'ungguli', 'yu', 'pasangan', 'terbuka', 'jalannya', 'indonesia', 'poliinitya', 'game', 'final', 'tiongkok', 'greysia', 'poinini', 'greysianitya', 'maheswari'

3.1.3.6 Analisis *Feature Extraction* dengan TF-IDF

Dalam *information retrieval* atau *text mining, term frequency – inverse document frequency* (TF-IDF) merupakan suatu cara untuk memberikan bobot hubungan suatu kata (term) terhadap dokumen. Dimana *term* adalah kata hasil dari proses *preprocessing* stopword. *tf* menyatakan jumlah berapa banyak keberadaan suatu *term* (t) dalam satu dokumen (d), df merupakan perhitungan berapa

banyaknya dokumen dimana suatu term (t) muncul dan *idf* berfungsi mengurangi bobot suatu *term* jika kemunculannya banyak tersebar di seluruh koleksi dokumen. Hal ini dikarenakan terkadang suatu *term* muncul di hampir sebagian besar dokumen mengakibatkan proses pencarian *term* unik terganggu.

Untuk mendapatkan nilai *idf* dilakukan perhitungan dengan persamaan (2.1). Setelah mendapatkan nilai *tf* dan *idf* maka selanjutnya menghitung pembobotan pada *term*, untuk mendapatkan bobot (W) masing-masing dokumen pada setiap *term* dilakukan persamaan (2.2). Berikut alur proses TF-IDF pada Gambar 3.9



Gambar 3.9 *Flowchart Feature Extraction* TF-IDF

Tabel 3.7 merupakan contoh dokumen yang sudah melalui proses *preprocessing*. Doc1 kelas ekonomi, Doc2 kelas kesehatan , Doc3 kelas teknologi, Doc4 kelas pariwisata dan Doc5 kelas olahraga yang merupakan contoh artikel data input pada analisis masukan,sebagai berikut.

Tabel 3.7 Term dari Data Artikel

Dokumen Artikel	**Kumpulan Term**	**Kelas**
Doc1	persen leasing penjualan barron jatuh pesawat keuangan saham boeing laporan maskapai perusahaan co lemahnya penerbangan minggu akibat memilih publikasi dipaparkan permintaan melambat menunda order	ekonomi
Doc2	sulit gejala penderitanya pikiran gangguan mendampingi lakukan orang seringkali pasien perubahan jangka berinteraksi mengakibatkan kadang kebanyakan motorik bicara usia demensia memori pendek diderita merawat terpengaruh demesia berpikir kumpulan kemampuan	kesehatan
Doc3	pesan muncul ios kendala pengguna mengeluhkan sore kemarin mengakses instan kamis menemui rabu keluhan nampaknya platform pantauan layanan whatsapp iphone	teknologi
Doc4	pemerintah kabupaten memamerkan menggelar pertanian menetapkan agro banyuwangi panggung suguhan acara event bupati pangan unggulannya taman abdullah menampilkan azwar mewajibkan expo holtikultura petani pns kebijakan perhatian bentuk blambangan senin tanaman buah anas	pariwisata

	melibatkan asli lokal produk kompastravel proteksi	
Doc5	series memimpin india yuanting lawan kemenangan menumbangkan pertandingan menit puteri yangtang sukses super yutang langsung menang gim mementahkan perolehan perempat berhasil ganda serangan krishinda ungguli yu pasangan terbuka jalannya indonesia poliinitya game final tiongkok greysia poin ini greysianitya maheswari	olahraga

Proses awal dilakukan perhitungan kata (*term*) pada tiap dokumen, sehingga mendapatkan frekuensi *term* (*tf*). Selanjutnya dilakukan *document frequency* (*df*).

Setelah didapatkan hasil perhitungan *df*, selanjutnya dilakukan perhitungan *inverse document frequency* (*idf*) dengan persamaan (2.1) sebagai berikut:

$$idf\,(t) = \mathrm{Log}\left(\frac{D}{df}\right) + 1 \qquad (2.1)$$

Diambil contoh pada kata "acara". Diperoleh banyaknya dokumen (D) = 5 dan *df* pada kata "acara" = 1. Maka, perhitungannya sebagai berikut:

$$idf\,(t) = \mathrm{Log}\left(\frac{5}{1}\right)+1 = 1.69897$$

Untuk sebagian hasil perhitungan *tf*, *df* dan *idf* dapat dilihat pada Tabel 3.8, sedangkan untuk hasil selengkapnya dapat dilihat pada Lampiran B Tabel B.1.

Selanjutnya untuk mendapatkan bobot term, maka dilakukan perhitungan *tf* - *idf* dengan persamaan 2.2 sebagai berikut:

$$tf\text{-}idf\,(t) = tf\,(t,d) * idf\,(t) \qquad (2.2)$$

Diperoleh *tf* (t,d) = 1 dengan *idf* (t) = 1.69897. Maka, perhitungannya sebagai berikut:

tf-idf (t) = 1 * 1.69897 = 1.69897

Sehingga nilai bobot untuk kata "acara" adalah 1.69897.

Untuk sebagian hasil dari perhitungan pembobotan *tf-idf* dapat dilihat pada Tabel 3.9, sedangkan untuk hasil selengkapnya dapat dilihat pada Lampiran D Tabel D.1.

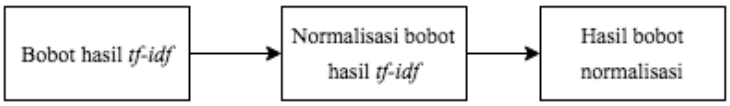
Tabel 3.8 Hasil Perhitungan *tf, df* dan *idf*

Term	*tf*					*df*	*idf*
	Doc1	Doc2	Doc3	Doc4	Doc5		
abdullah	0	0	0	1	0	1	1.69897
acara	0	0	0	1	0	1	1.69897
agro	0	0	0	1	0	1	1.69897
akibat	1	0	0	0	0	1	1.69897
anas	0	0	0	2	0	1	1.69897
asli	0	0	0	1	0	1	1.69897
azwar	0	0	0	1	0	1	1.69897
banyuwangi	0	0	0	4	0	1	1.69897
barron	1	0	0	0	0	1	1.69897
bentuk	0	0	0	1	0	1	1.69897
berhasil	0	0	0	0	1	1	1.69897
berinteraksi	0	1	0	0	0	1	1.69897
berpikir	0	1	0	0	0	1	1.69897
bicara	0	1	0	0	0	1	1.69897
blambangan	0	0	0	1	0	1	1.69897
boeing	1	0	0	0	0	1	1.69897
buah	0	0	0	3	0	1	1.69897
bupati	0	0	0	1	0	1	1.69897
co	1	0	0	0	0	1	1.69897
demensia	0	3	0	0	0	1	1.69897
diderita	0	1	0	0	0	1	1.69897
dipaparkan	1	0	0	0	0	1	1.69897
event	0	0	0	1	0	1	1.69897
expo	0	0	0	1	0	1	1.69897
final	0	0	0	0	1	1	1.69897
game	0	0	0	0	1	1	1.69897
ganda	0	0	0	0	1	1	1.69897
gangguan	0	1	0	0	0	1	1.69897
gejala	0	1	0	0	0	1	1.69897
gim	0	0	0	0	2	1	1.69897

Tabel 3.9 Hasil Perhitungan *tf-idf*

Term	*tf-idf (t)*				
	Doc1	**Doc2**	**Doc3**	**Doc4**	**Doc5**
abdullah	0	0	0	1.69897	0
acara	0	0	0	1.69897	0
agro	0	0	0	1.69897	0
akibat	1.69897	0	0	0	0
anas	0	0	0	3.39794	0
asli	0	0	0	1.69897	0
azwar	0	0	0	1.69897	0
banyuwangi	0	0	0	6.79588	0
barron	1.69897	0	0	0	0
bentuk	0	0	0	1.69897	0
berhasil	0	0	0	0	1.69897
berinteraksi	0	1.69897	0	0	0
berpikir	0	1.69897	0	0	0
bicara	0	1.69897	0	0	0
blambangan	0	0	0	0	0
boeing	1.69897	0	0	1.69897	0
buah	0	0	0	5.09691	0
bupati	0	0	0	1.69897	0
co	1.69897	0	0	0	0
demensia	0	5.09691	0	0	0
diderita	0	1.69897	0	0	0
dipaparkan	1.69897	0	0	0	0
event	0	0	0	1.69897	0
expo	0	0	0	1.69897	0
final	0	0	0	0	1.69897
game	0	0	0	0	1.69897
ganda	0	0	0	0	1.69897
gangguan	0	1.69897	0	0	0
gejala	0	1.69897	0	0	0
gim	0	0	0	0	3.39794

Persamaan 2.1 dapat dinormalisasikan dengan Persamaan 2.2. Normalisasi ini dilakukan dengan tujuan untuk menstandarisasi nilai bobot ke dalam interval 0 sampai dengan 1. Berikut proses normalisasi dapat dilihat pada gambar 3.10.



Gambar 3.10 Proses Normalisasi

Diambil contoh bobot normalisasi pada kata "acara" dengan nilai bobot *tf-idf* 1.69897 dari Doc4 dan Perhitungan normalisasi dilakukan dengan rumus sebagai berikut.

$$w_{ij} = \frac{tf*(\log\left(\frac{D}{df}\right)+1)}{\sqrt{\sum_{k=1}^{t}(tf_{ik})^2*\left[(\log\left(\frac{D}{df}\right))+1\right]^2}} \quad (2.3)$$

$$w_{(acara)} = \frac{1.69897}{\sqrt{\begin{matrix}1.69897^2+1.69897^2+1.69897^2+0^2+3.39794^2+1.69897^2+1.69897^2+\\6.79588^2+0^2+1.69897^2+0^2+0^2+0^2+0^2+0^2+1.69897^2+5.0961^2+\\0^2+0^2+0^2+0^2+1.69897^2+1.69897^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+\\0^2+1.69897^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+5.0961^2+0^2+0^2+0^2+\\1.69897^2+0^2+0^2+0^2+0^2+0^2+0^2+1.69897^2+1.69897^2+1.69897^2\end{matrix}}} = 0.09950372$$

Untuk sebagian hasil dari perhitungan normalisasi w_{ij} dapat dilihat pada Tabel 3.10, sedangkan untuk hasil selengkapnya dapat dilihat pada Lampiran E.

Tabel 3.10 Hasil Normalisasi

Term	w_{ij}				
	Doc1	**Doc2**	**Doc3**	**Doc4**	**Doc5**
abdullah	0	0	0	0.09950372	0
acara	0	0	0	0.09950372	0
agro	0	0	0	0.09950372	0
akibat	0.16903085	0	0	0	0
anas	0	0	0	0.19900744	0
asli	0	0	0	0.09950372	0
azwar	0	0	0	0.09950372	0
banyuwangi	0	0	0	0.39801488	0
barron	0.16903085	0	0	0	0
bentuk	0	0	0	0.09950372	0
berhasil	0	0	0	0	0.15075567
berinteraksi	0	0.14002801	0	0	0
berpikir	0	0.14002801	0	0	0
bicara	0	0.14002801	0	0	0

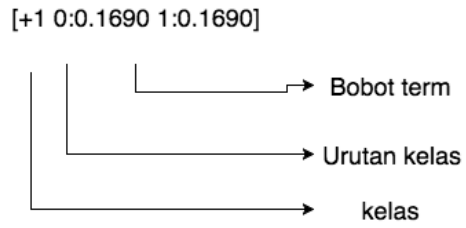
blambangan	0	0	0	0	0
boeing	0.16903085	0	0	0.09950372	0
buah	0	0	0	0.09950372	0
bupati	0	0	0	0.09950372	0
co	0.16903085	0	0	0	0
demensia	0	0.42008403	0	0	0
diderita	0	0.14002801	0	0	0
dipaparkan	0.16903085	0	0	0	0
event	0	0	0	1.69897	0
expo	0	0	0	1.69897	0
final	0	0	0	0	0.15075567
game	0	0	0	0	0.15075567
ganda	0	0	0	0	0.15075567
gangguan	0	0.14002801	0	0	0
gejala	0	0.14002801	0	0	0
gim	0	0	0	0	0.30151134

Feature extraction dengan pembobotan Tf-Idf merupakan suatu cara untuk memberikan bobot hubungan suatu kata (*term*) terhadap dokumen. Hasil dari nilai bobot tersebut kemudian akan digunakan dalam pembentukan vektor pada klasifikasi SVM. Data hasil w_{ij} akan diubah kedalam bentuk matriks untuk digunakan dalam representasi data.

3.1.4 Analisis Klasifikasi menggunakan Metode SVM

3.1.4.1 Representasi Data

Dalam penelitian ini untuk representasi data menggunakan format *sparse data representation* dengan vector sebagai inputnya. Format data input untuk klasifikasi SVM dalam penelitian ini adalah [+1 0:0.1690 1:0.1690]. Dengan masukan pertama +1 atau -1 menyatakan dua kelas. Angka kedua menyatakan dimensi (row_id) dan angka ketiga setelah tanda ":" menyatakan bobot dari *term*, setiap *term* didalam dokumen dipisahkan dengan spasi, [*label* *row-id*:*bobot*$_{(row\text{-}id)}$ *row-id*$_n$:*bobot*$_{(row\text{-}idn)}$].



Data *training* dan data uji akan diubah menjadi data vektor. Data *training* adalah data latih artikel yang sudah memiliki kelas. Sedangkan data uji adalah data uji artikel yang belum memiliki kelas. Data training digunakan untuk melakukan proses pembelajaran terhadap sistem. Proses pembelajaran ini akan menghasilkan model baru yang akan digunakan pada klasifikasi.

Sebagai contoh dalam pengubahan data teks menjadi data vektor dari hasil bobot normalisasi yang diambil dari contoh kasus yang telah melalui tahap pembobotan TF-IDF pada Tabel 3.9. Dokumen yang digunakan adalah Doc1, Doc2, Doc3, Doc4 dan Doc5, sebagai inisialisasi Doc1 diberi label +1 selain itu diberi label -1 maka Doc2, Doc3, Doc4 dan Doc5 diberi label -1. Untuk contoh hasil data artikel menjadi format data vektor dapat dilihat pada Tabel 3.11.

Tabel 3.11 Data Artikel menjadi format data vektor

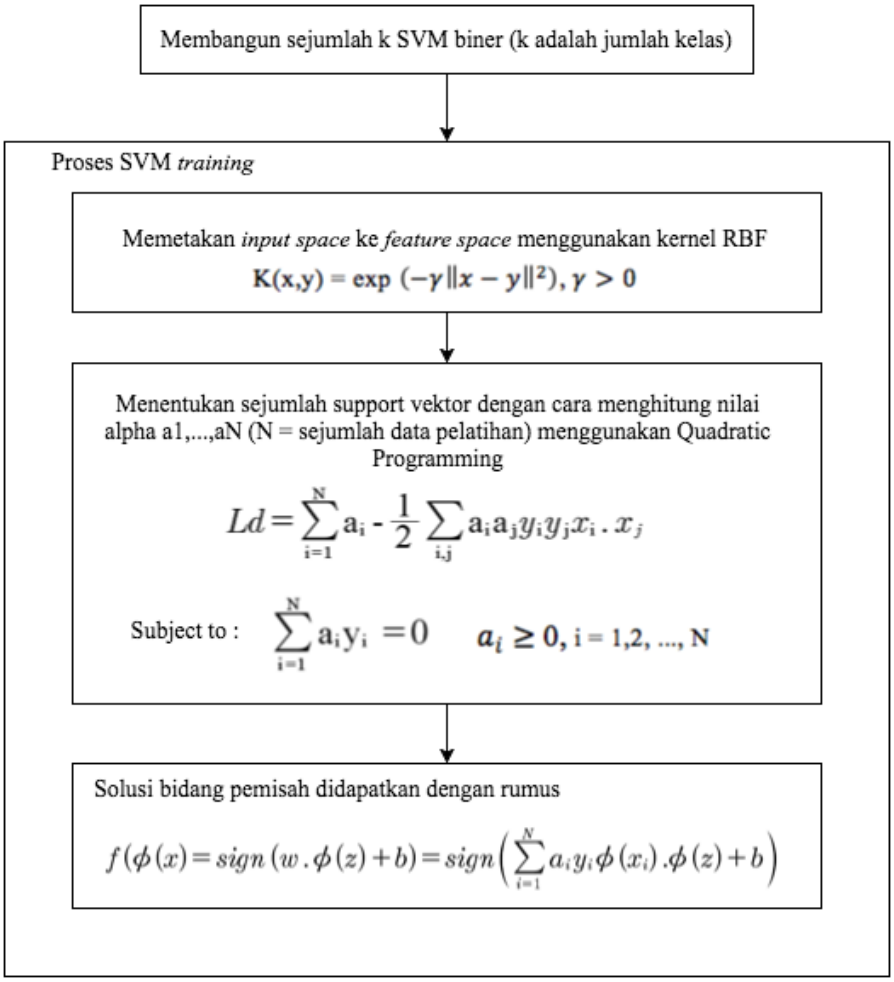
Doc1	Doc2	Doc3	Doc4	Doc5
[+1 1:0	[-1 1:0	[-1 1:0	[-1 1:0.09950372	[-1 1:0
2:0	2:0	2:0	2:0.09950372	2:0
3:0	3:0	3:0	3:0.09950372	3:0
4:0.16903085	4:0	4:0	4:0	4:0
5:0	5:0	5:0	5:0.19900744	5:0
6:0	6:0	6:0	6:0.09950372	6:0
7:0	7:0	7:0	7:0.09950372	7:0
8:0	8:0	8:0	8:0.39801488	8:0
9:0.16903085	9:0	9:0	9:0	9:0
10:0	10:0	10:0	10:0.09950372	10:0
11:0	11:0	11:0	11:0	11:0.15075567
12:0	12:0.14002801	12:0	12:0	12:0
13:0	13:0.14002801	13:0	13:0	13:0
14:0	14:0.14002801	14:0	14:0	14:0
15:0	15:0	15:0	15:0	15:0
16:0.16903085	16:0	16:0	16:0.09950372	16:0

17:0	17:0	17:0	17:0.09950372	17:0
18:0	18:0	18:0	18:0.09950372	18:0
19:0.16903085	19:0	19:0	19:0	19:0
20:0	20:0.42008403	20:0	20:0	20:0
21:0	21:0.14002801	21:0	21:0	21:0
22:0.16903085	22:0	22:0	22:0	22:0
23:0	23:0	23:0	23: 0.09950372	23:0
24:0	24:0	24:0	24: 0.09950372	24:0
25:0	25:0	25:0	25:0	25:0.15075567
26:0	26:0	26:0	26:0	26:0.15075567
27:0	27:0	27:0	27:0	27:0.15075567
28:0	28:0.14002801	28:0	28:0	28:0
29:0	29:0.14002801	29:0	29:0	29:0
30:0]	30:0]	30:0]	30:0]	30:0.30151134]

Untuk kata atau *term* yang muncul lebih dari sekali dalam sebuah kumpulan *term* akan diwakili oleh sebuah data vektor saja dengan bobot normalisasi yang bersesuaian. Data vektor pada sebuah kalimat diurutkan berdasarkan indeks kata dari terkecil hingga terbesar. Tahap ini akan terus berulang hingga semua *term* pada data artikel sudah berubah ke format data vektor.

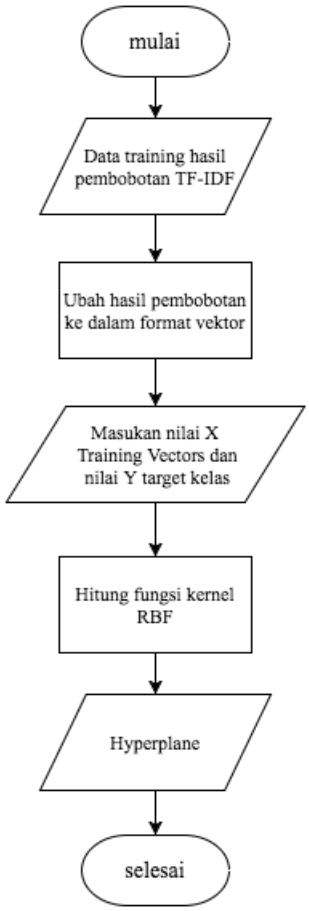
3.1.4.2 Analisis Proses SVM Training

Pada proses pelatihan SVM bertujuan untuk menemukan vektor α dan konstanta b yang akan digunakan dalam menemukan fungsi pemisah (Hyperplane) dengan margin yang maksimal. Vektor $\alpha > 0$ dinamakan support vector dan menyatakan data training yang diperlukan untuk mewakili fungsi keputusan yang optimal [12]. Konstanta b menentukan lokasi fungsi pemisah relative terhadap titik asal (origin)[12]. Dalam proses pelatihan dibutuhkan satu set input-output data atau dalam kasus ini akan digunakan dokumen yang memiliki kelas ekonomi, kelas kesehatan, kelas teknologi, kelas pariwisata dan kelas olahraga. Pada penelitian ini menggunakan SVM non linear, maka perlu menggunakan fungsi kernel atau kernel *trick*. Fungsi kernel yang digunakan dalam penelitian ini adalah fungsi kernel RBF. Berikut adalah blok diagram proses pelatihan menggunakan SVM



Gambar 3.11 Blok Diagram Proses *Multiclass* SVM *Training*

Berikut alur proses *Multiclass* SVM training dapat dilihat pada Gambar 3.12



Gambar 3.12 *Flowchart* SVM *Training*

Berikut adalah gambaran dari sebagian data vektor artikel yang sudah memiliki kelas yang disajikan dalam bentuk tabel yang akan digunakan dalam analisis training. Untuk data lengkap dapat dilihat di Lampiran F.

Tabel 3.12 Data *Support* Vektor Artikel

	Doc1	**Doc2**	**Doc3**	**Doc4**	**Doc5**
term1	0	0	0	0.09950372	0
term2	0	0	0	0.09950372	0
term3	0	0	0	0.09950372	0
term4	0.16903085	0	0	0	0
term5	0	0	0	0.19900744	0
term6	0	0	0	0.09950372	0
term7	0	0	0	0.09950372	0
term8	0	0	0	0.39801488	0
term9	0.16903085	0	0	0	0
term10	0	0	0	0.09950372	0

term11	0	0	0	0	0.15075567
term12	0	0.14002801	0	0	0
term13	0	0.14002801	0	0	0
term14	0	0.14002801	0	0	0
term15	0	0	0	0.09950372	0
term16	0.16903085	0	0	0	0
term17	0	0	0	0.29851116	0
term18	0	0	0	0.09950372	0
term19	0.16903085	0	0	0	0
term20	0	0.42008403	0	0	0
term21	0	0.14002801	0	0	0
term22	0.16903085	0	0	0	0
term23	0	0	0	0.09950372	0
term24	0	0	0	0.09950372	0
term25	0	0	0	0	0.15075567
term26	0	0	0	0	0.15075567
term27	0	0	0	0	0.15075567
term28	0	0.14002801	0	0	0
term29	0	0.14002801	0	0	0
term30	0	0	0	0	0.30151134
term31	0	0	0	0	0.15075567
term32	0	0	0	0	0.15075567
term33	0	0	0	0.09950372	0
term34	0	0	0	0	0.15075567
term35	0	0	0	0	0.15075567
term36	0	0	0.1767767	0	0
term37	0	0	0.1767767	0	0
term38	0	0	0.1767767	0	0
term39	0	0	0	0	0.15075567
term40	0	0.14002801	0	0	0
term41	0.16903085	0	0	0	0
term42	0	0	0	0.29851116	0
term43	0	0.14002801	0	0	0
term44	0	0	0.1767767	0	0
term45	0	0.14002801	0	0	0
term46	0	0	0	0.09950372	0
term47	0	0	0.1767767	0	0
term48	0	0.14002801	0	0	0
term49	0	0	0.1767767	0	0
term50	0	0	0	0	0.15075567
term51	0	0	0.1767767	0	0
term52	0.16903085	0	0	0	0
term53	0	0	0	0.09950372	0
term54	0	0	0	0	0.15075567
term55	0	0.14002801	0	0	0

term56	0	0.14002801	0	0	0
term57	0	0	0	0	0.15075567
term58	0.16903085	0	0	0	0
term59	0	0	0	0	0.15075567
term60	0	0	2	0	0

3.1.4.2.1 Analisis Training *one vs all*

Dengan metode ini, untuk masalah klasifikasi k-kelas, ditemukan k fungsi pemisah dimana k adalah banyaknya kelas. Misalkan ada fungsi pemisah dengan nama ρ. Dalam metode ini, ρ^i ditrain dengan semua data dari kelas ke-I dengan label +1 dan semua data dari kelas lain dengan label -1. Dalam tahap ini diambil contoh kasus yang telah melalui tahap *preprocessing* dan pembobotan TF-IDF. Dokumen yang digunakan adalah Doc1, Doc2 , Doc3, Doc4 dan Doc5 dengan sudah diberikan label, dimana Doc1 diberikan label +1 sedangkan Doc2 , Doc3, Doc4 dan Doc5 diberikan label -1.

Tahap selanjutnya adalah menghitung nilai x. Nilai x pada tabel akan digunakan untuk perhitungan kernel atau perhitungan dot product. Nilai x didapat dari nilai term pada dokumen. Nilai pada kolom Doc1 merupakan nilai x1, nilai pada kolom Doc2 merupakan nilai x2 , nilai pada kolom Doc3 merupakan nilai x3, nilai pada kolom Doc4 merupakan nilai x4 dan nilai pada kolom Doc5 merupakan nilai x5. Pada tahap selanjutnya yaitu melakukan kernelisasi menggunakan fungsi kernel RBF yang didefinisikan sebagai K(x,y) = exp $(-\gamma\|x-y\|^2)$, $\gamma > 0$

Formulasi yang digunakan adalah dualitas Lagrange multiplier (persamaan 2.23) sehingga untuk w harus menggunakan persamaan 2.17 yang sudah dimodifikasi untuk x dengan fungsi kernelnya, menjadi:

W= $\sum_{i=1}^{N} a_i y_i \emptyset x_i$

Selanjutnya harus dilakukan kernelisasi pada set data dari fitur dimensi lama sehingga mendapatkan set data dengan fitur baru dimensi tinggi. Dengan kernel $K(x,x_i)$ = exp $(-\gamma\|x-x_i\|^2)$, dan set data berdimensi Nx1 maka akan didapatkan dimensi baru NxN, dimana N adalah banyaknya data. Dengan menggunakan informasi Tabel 3.11, maka untuk K(1,1), K(1,2), …, K(5,5) dilakukan perhitungan dengan kernel pada semua data dengan $\gamma = 0.5$. Dimana K adalah matriks kernel.

Pada Tabel 3.13 menampilkan sebagian dari perhitungan pengurangan x-x_i. Untuk hasil lengkap perhitungan pengurangan x-x_i dapat dilihat di Lampiran G.

Tabel 3.13 Perhitungan x-x_i untuk K(1,1), K(1,2) , K(1,3), K(1,4) dan K(1,5)

x1-x1	x1-x2	x1-x3	x1-x4	x1-x5
0	0	0	-0.09950372	0
0	0	0	-0.09950372	0
0	0	0	-0.09950372	0
0	0.16903085	0.16903085	0.16903085	0.16903085
0	0	0	-0.19900744	0
0	0	0	-0.09950372	0
0	0	0	-0.09950372	0
0	0	0	-0.39801488	0
0	0.16903085	0.16903085	0.16903085	0.16903085
0	0	0	-0.09950372	0
0	0	0	0	-0.15075567
0	-0.14002801	0	0	0
0	-0.14002801	0	0	0
0	-0.14002801	0	0	0
0	0	0	-0.09950372	0
0	0.16903085	0.16903085	0.16903085	0.16903085
0	0	0	-0.29851116	0
0	0	0	-0.09950372	0
0	0.16903085	0.16903085	0.16903085	0.16903085
0	-0.42008403	0	0	0
0	-0.14002801	0	0	0
0	0.16903085	0.16903085	0.16903085	0.16903085
0	0	0	-0.09950372	0
0	0	0	-0.09950372	0
0	0	0	0	-0.15075567
0	0	0	0	-0.15075567
0	0	0	0	-0.15075567
0	-0.14002801	0	0	0
0	-0.14002801	0	0	0
0	0	0	0	-0.30151134
0	0	0	0	-0.15075567
0	0	0	0	-0.15075567
0	0	0	-0.09950372	0

0	0	0	0	-0.15075567
0	0	0	0	-0.15075567
0	0	-0.1767767	0	0
0	0	-0.1767767	0	0
0	0	-0.1767767	0	0
0	0	0	0	-0.15075567
0	-0.14002801	0	0	0
0	0.16903085	0.16903085	0.16903085	0.16903085
0	0	0	-0.29851116	0
0	-0.14002801	0	0	0
0	0	-0.1767767	0	0
0	-0.14002801	0	0	0
0	0	0	-0.09950372	0
0	0	-0.1767767	0	0
0	-0.14002801	0	0	0
0	0	-0.1767767	0	0
0	0	0	0	-0.15075567
0	0	-0.1767767	0	0
0	0.16903085	0.16903085	0.16903085	0.16903085
0	0	0	-0.09950372	0
0	0	0	0	-0.15075567
0	-0.14002801	0	0	0
0	-0.14002801	0	0	0
0	0	0	0	-0.15075567
0	0.16903085	0.16903085	0.16903085	0.16903085
0	0	0	0	-0.15075567
0	0	-0.3535534	0	0

Setelah didapatkan hasil pengurangan x-x_i, maka selanjutnya dilakukan perhitungan untuk mendapatkan panjang vektor yang digambarkan pada Tabel 3.14. Besar atau panjang vektor misalnya r (x,y,z) dapat ditentukan dengan rumus.

$$|r| = \sqrt{x^2 + y^2 + z^2}$$

Tabel 3.14 Perhitungan panjang vektor untuk K(1,1), K(1,2) , K(1,3), K(1,4) dan K(1,5)

||x1-x1|| =	$\sqrt{\begin{array}{l}0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\end{array}}$	0
||x1-x2|| =	$\sqrt{\begin{array}{l}0^2+0^2+0^2+0.16903085^2+0^2+0^2+0^2+0^2+0.16903085^2+0^2+\\0^2-0.14002801^2-0.14002801^2-0.14002801^2+0^2+0.16903085^2+0^2\\+0^2+0.16903085^2-0.42008403^2-0.14002801^2+0.16903085^2+0^2+0^2\\+0^2+0^2+0^2-0.14002801^2-0.14002801^2+0^2+\\0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2-0.14002801^2+\\0.16903085^2+0^2-0.14002801^2+0^2-0.14002801^2+0^2+0^2-0.14002801^2\\+0^2+0^2+0^2+0.16903085^2+0^2+0^2-0.14002801^2-0.14002801^2+0^2+0.16903085^2+0^2\end{array}}$	1.41421
||x1-x3|| =	$\sqrt{\begin{array}{l}0.16903085^2+0.16903085^2+0^2+0^2+0^2\\+0.16903085^2+0.16903085^2+0^2+0^2+0.16903085^2\\+0^2+0^2+(-0.1767767)^2+(-0.1767767)^2+(-0.1767767)^2+0^2\\+0.16903085^2+0^2+(-0.1767767)^2+0^2+(-0.1767767)^2+0^2\\+(-0.1767767)^2+(-0.1767767)^2+0.16903085^2+0^2+0^2\\+0.16903085^2+(-0.35355339)^2+0.16903085^2+0.16903085^2\\+0.16903085^2+0.16903085^2+0.16903085^2+0^2+0^2+(-0.1767767)^2\\+0^2+(-0.1767767)^2+(-0.1767767)^2+0.16903085^2+0^2\\+0.16903085^2+0^2+(-0.1767767)^2+(-0.1767767)^2+0^2\\+0.16903085^2+(-0.1767767)^2+0.16903085^2+(-0.35355339)^2\\+0.16903085^2+0.16903085^2+0.16903085^2+0^2+0.16903085^2\\+(-0.1767767)^2+0.50709255^2+0^2+(-0.35355339)^2\\+0.3380617^2+(-0.1767767)^2+0.16903085^2+0^2\\+(-0.1767767)^2+0^2+0^2+0^2+(-0.35355339)^2\end{array}}$	1.41421

||x1-x4|| =	$\sqrt{\begin{array}{l}0.16903085^2+0.16903085^2+0^2+0^2+0^2\\+0.16903085^2+0.16903085^2+0^2+0^2+0.16903085^2\\+0^2+0^2+(-0.1767767)^2+(-0.1767767)^2+(-0.1767767)^2+0^2\\+0.16903085^2+0^2+(-0.1767767)^2+0^2+(-0.1767767)^2+0^2\\+(-0.1767767)^2+(-0.1767767)^2+0.16903085^2+0^2+0^2\\+0.16903085^2+(-0.35355339)^2+0.16903085^2+0.16903085^2\\+0.16903085^2+0.16903085^2+0.16903085^2+0^2+0^2+(-0.1767767)^2\\+0^2+(-0.1767767)^2+(-0.1767767)^2+0.16903085^2+0^2\\+0.16903085^2+0^2+(-0.1767767)^2+(-0.1767767)^2+0^2\\+0.16903085^2+(-0.1767767)^2+0.16903085^2+(-0.35355339)^2\\+0.16903085^2+0.16903085^2+0.16903085^2+0^2+0.16903085^2\\+(-0.1767767)^2+0.50709255^2+0^2+(-0.35355339)^2\\+0.3380617^2+(-0.1767767)^2+0.16903085^2+0^2\\+(-0.1767767)^2+0^2+0^2+0^2+(-0.35355339)^2\end{array}}$	1.41421
||x1-x5|| =	$\sqrt{\begin{array}{l}0.16903085^2+0.16903085^2+0^2+0^2+0^2\\+0.16903085^2+0.16903085^2+0^2+0^2+0.16903085^2\\+0^2+0^2+(-0.1767767)^2+(-0.1767767)^2+(-0.1767767)^2+0^2\\+0.16903085^2+0^2+(-0.1767767)^2+0^2+(-0.1767767)^2+0^2\\+(-0.1767767)^2+(-0.1767767)^2+0.16903085^2+0^2+0^2\\+0.16903085^2+(-0.35355339)^2+0.16903085^2+0.16903085^2\\+0.16903085^2+0.16903085^2+0.16903085^2+0^2+0^2+(-0.1767767)^2\\+0^2+(-0.1767767)^2+(-0.1767767)^2+0.16903085^2+0^2\\+0.16903085^2+0^2+(-0.1767767)^2+(-0.1767767)^2+0^2\\+0.16903085^2+(-0.1767767)^2+0.16903085^2+(-0.35355339)^2\\+0.16903085^2+0.16903085^2+0.16903085^2+0^2+0.16903085^2\\+(-0.1767767)^2+0.50709255^2+0^2+(-0.35355339)^2\\+0.3380617^2+(-0.1767767)^2+0.16903085^2+0^2\\+(-0.1767767)^2+0^2+0^2+0^2+(-0.35355339)^2\end{array}}$	1.41421

Hasil dari panjang vektor pada Tabel 3.14 kemudian dimasukan ke dalam persamaan exp $(-\gamma\|x-x_i\|^2)$ sebagai berikut:

K(1,1) = exp $(-\gamma\|x1-x1\|^2)$
=exp (-0.5(0^2))
=exp (0)
=1

K(1,2) = exp $(-\gamma\|x1-x2\|^2)$
=exp (-0.5(1.41421^2))
=exp (-1)
=0.3679

K(1,3) = exp $(-\gamma\|x1-x3\|^2)$
=exp (-0.5(1.41421^2))
=exp (-1)
=0.3679

K(1,4) = exp $(-\gamma\|x1-x3\|^2)$
=exp (-0.5(1.41421^2))
=exp (-1)
=0.3679

K(1,5) = exp $(-\gamma\|x1 - x3\|^2)$

=exp (-0.5(1.41421^2))

=exp (-1)

=0.3679

Untuk hasil perhitungan nilai exp $(-\gamma\|x - x_i\|^2)$ selanjutnya didapatkan nilai pada Tabel 3.15 .

Tabel 3.15 Hasil perhitungan nilai K dengan kernel RBF

K(1,1)	**K(1,2)**	**K(1,3)**	**K(1,4)**	$\boldsymbol{K(1,5)}$
1	0.367879441	0.367879437	0.36787944	0.367879449
K(2,1)	**K(2,2)**	**K(2,3)**	**K(2,4)**	**K(2,5)**
0.367879441	1	0.367879433	0.367879435	0.367879445
$\mathbf{K(3,1)}$	**K(3,2)**	**K(3,3)**	**K(3,4)**	**K(3,5)**
0.367879437	0.367879433	1	0.367879432	0.367879441
K(4,1)	**K(4,2)**	**K(4,3)**	**K(4,4)**	**K(4,5)**
0.36787944	0.367879435	0.367879432	1	0.367879443
K(5,1)	**K(5,2)**	**K(5,3)**	**K(5,4)**	**K(5,5)**
0.367879449	0.367879445	0.367879441	0.367879443	1

Setelah dilakukan perhitungan, maka matriks kernel K yang terbentuk adalah sebagai berikut.

$$K = \begin{bmatrix} \mathrm{K}(1,1) & \mathrm{K}(2,1) & \mathrm{K}(3,1) & \mathrm{K}(4,1) & \mathrm{K}(5,1) \\ \mathrm{K}(1,2) & \mathrm{K}(2,2) & \mathrm{K}(3,2) & \mathrm{K}(4,2) & \mathrm{K}(5,2) \\ \mathrm{K}(1,3) & \mathrm{K}(2,3) & \mathrm{K}(3,3) & \mathrm{K}(4,3) & \mathrm{K}(5,3) \\ \mathrm{K}(1,4) & \mathrm{K}(2,4) & \mathrm{K}(3,4) & \mathrm{K}(4,4) & \mathrm{K}(5,4) \\ \mathrm{K}(1,5) & \mathrm{K}(2,5) & \mathrm{K}(3,5) & \mathrm{K}(4,5) & \mathrm{K}(5,5) \end{bmatrix}$$

$$K = \begin{bmatrix} 1 & 0.367879441 & 0.367879437 & 0.36787944 & 0.367879449 \\ 0.367879441 & 1 & 0.367879433 & 0.367879435 & 0.367879445 \\ 0.367879437 & 0.367879433 & 1 & 0.367879432 & 0.367879441 \\ 0.36787944 & 0.367879435 & 0.367879432 & 1 & 0.367879443 \\ 0.367879449 & 0.367879445 & 0.367879441 & 0.367879443 & 1 \end{bmatrix}$$

Kemudian tahap selanjutnya adalah melakukan perhitungan terhadap *y*. Untuk nilai y adalah nilai dari label atau nilai dari kelas yang telah diberikan. Nilai y tersebut dapat dilihat pada tabel 3.16.

Tabel 3. 16 Nilai label pada y

y_1	y_2	y_3	y_4	y_5
1	-1	-1	-1	-1

Selanjutnya dilakukan perhitungan $y_i y_j$ sebanyak N data. Hasil dari perhitungan tersebut dapat dilihat pada Tabel 3.17.

Tabel 3.17 Hasil perhitungan $y_i y_j$

$\boldsymbol{y_1 y_1}$	$\boldsymbol{y_1 y_2}$	$\boldsymbol{y_1 y_3}$	$\boldsymbol{y_1 y_4}$	$\boldsymbol{y_1 y_5}$
1	-1	-1	-1	-1
$\boldsymbol{y_2 y_1}$	$\boldsymbol{y_2 y_2}$	$\boldsymbol{y_2 y_3}$	$\boldsymbol{y_2 y_4}$	$\boldsymbol{y_2 y_5}$
-1	1	1	1	1
$\boldsymbol{y_3 y_1}$	$\boldsymbol{y_3 y_2}$	$\boldsymbol{y_3 y_3}$	$\boldsymbol{y_3 y_4}$	$\boldsymbol{y_3 y_5}$
-1	1	1	1	1
$\boldsymbol{y_4 y_1}$	$\boldsymbol{y_4 y_2}$	$\boldsymbol{y_4 y_3}$	$\boldsymbol{y_4 y_4}$	$\boldsymbol{y_4 y_5}$
-1	1	1	1	1
$\boldsymbol{y_5 y_1}$	$\boldsymbol{y_5 y_2}$	$\boldsymbol{y_5 y_3}$	$\boldsymbol{y_5 y_4}$	$\boldsymbol{y_5 y_5}$
-1	1	1	1	1

Sehingga matriks yang terbentuk dari perhitungan $y_i y_j$ hasilnya sebagai berikut:

$$y_i y_j = \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Matriks K diatas setiap elemenya merupakan hasil exp $(-\gamma\|x - x_i\|^2)$, yang akan berkorelasi dengan $a_i a_j$ dalam persamaan 2.23. Dengan menggunakan kernel K sebagai pengganti dot-product $x_i x_j$ dalam persamaan dualitas Lagrange multiplier 2.23, didapatkan:

Maksimalkan:

Ld max = $a_1 + a_2 + a_3 + a_4 + a_5$ - $\frac{1}{2}$ ($a_1{}^2$ + 0.73575888a_1a_2+ 0.73575888a_1a_3+0.73575888a_1a_4+0.73575888a_1a_5+ $a_2{}^2$ + 0.73575888a_2a_3+0.73575888a_2a_4 +0.73575888a_2a_5+ $a_3{}^2$ +0.73575888a_3a_4+0.73575888a_3a_5+ $a_4{}^2$ +0.73575888a_4a_5+ $a_5{}^2$))

Syarat 1: a_1 - a_2 - a_3 - a_4 - a_5= 0

Syarat 2: a_1 , a_2 , a_3, a_4, a_5 $\geq$ 0

Dalam fungsi tujuan, suku kedua sudah dikalikan dengan $y_i y_j$. Persamaan tersebut memenuhi standar *Quadaratic programming* sehingga bisa dibantu penyelesainnya dengan solver komersial untuk *Quadratic Programming* (QP). Dengan bantuan perangkat lunak, didapatkan hasil sebagai berikut.

a_1 = 0.25

a_2 = 0.25

a_3 = 0.25

a_4 = 0.25

a_5 = 1

b = -0.84196986

Hasil ini menunjukan bahwa semua data *training* adalah *support vector*. Karena nilai a > 0. Sementara nilai b didapatkan dari proses pelatihan yang dilakukan. Setelah menemukan semua a dan b, maka model SVM sudah siap digunakan untuk prediksi dengan menggunakan fungsi pemisah berikut [12`].

f(x) = w^T. x + b

dimana w^T. x + b = $a^T y$ $K(x_{training}, x_{testing})$+b

Untuk persamaan f(x) adalah sebagai berikut.

$$f(x) = a^T y \quad K(x_{training}, x_{testing}) + b$$

$$= \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \\ 1.0 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \end{bmatrix} \exp\left(-0.5 \left\| x_{training}\ x_{testing} \right\|^2\right) +$$

(-0.84196986)

3.1.4.2.2 Analisis Training *one vs one*

Dengan metode ini perlu menemukan k(k-1)/2 fungsi pemisah dimana setiap fungsi ditrain dengan data dari dua kelas. Dalam metode ini diambil contoh kasus yang telah melalui tahap pembobotan TF-IDF. Dokumen yang digunakan adalah Doc1 dan Doc2 pada Tabel 3.12 dengan sudah diberikan label, dimana Doc1 diberikan label 1 sedangkan Doc2 -1. Berikut gambaran label kelas data artikel yang direpresentasikan dalam bentuk Tabel 3.18.

Tabel 3.18 Artikel yang sudah memiliki kelas

	Doc1	**Doc2**
Y	1	-1

Nilai x pada tabel akan digunakan untuk perhitungan kernel. Nilai pada kolom Doc1 merupakan nilai x1 dan nilai pada kolom Doc2 merupakan nilai x2. Tahap selanjutnya yaitu melakukan kernelisasi menggunakan fungsi RBF yang didefinisikan sebagai K(x,y) = exp $(-\gamma\|x-y\|^2)$, $\gamma > 0$.

Selanjutnya dilakukan kernelisasi pada set data dari fitur dimensi lama sehingga mendapatkan set data dengan fitur baru dimensi tinggi. Dengan kernel $K(x,x_i)$ = exp $(-\gamma\|x-x_i\|^2)$, dan set data berdimensi Nx1 maka akan didaptakan dimensi baru NxN, dimana N adalah banyaknya data. Berikut merupakan contoh salah satu perhitungan kernel exp $(-\gamma\|x-x_i\|^2)$. Pada Tabel 3.19 menampilkan sebagian perhitungan pengurangan x-x_i pada perhitungan untuk K(1,1) , K(1,2) , K(2,1) dan K(2,2). Untuk hasil lengkap perhitungan pengurangan x-x_i dapat dilihat di Lampiran G.

Tabel 3.19 Perhitungan x-x_i untuk K(1,1), K(1,2), K(2,1) dan K(2,2)

x1-x1	x1-x2	x2-x1	x2-x2
0	0	0	0
0	0	0	0
0	0	0	0
0	0.16903085	-0.16903085	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0.16903085	-0.16903085	0
0	0	0	0
0	0	0	0
0	-0.14002801	0.14002801	0
0	-0.14002801	0.14002801	0
0	-0.14002801	0.14002801	0
0	0	0	0
0	0.16903085	-0.16903085	0
0	0	0	0
0	0	0	0
0	0.16903085	-0.16903085	0
0	-0.42008403	0.42008403	0
0	-0.14002801	0.14002801	0
0	0.16903085	-0.16903085	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	-0.14002801	0.14002801	0
0	-0.14002801	0.14002801	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0

0	0	0	0
0	0	0	0
0	0	0	0
0	-0.14002801	0.14002801	0
0	0.16903085	-0.16903085	0
0	0	0	0
0	-0.14002801	0.14002801	0
0	0	0	0
0	-0.14002801	0.14002801	0
0	0	0	0
0	0	0	0
0	-0.14002801	0.14002801	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0.16903085	-0.16903085	0
0	0	0	0
0	0	0	0
0	-0.14002801	0.14002801	0
0	-0.14002801	0.14002801	0
0	0	0	0
0	0.16903085	-0.16903085	0
0	0	0	0
0	0	0	0

Setelah didapatkan hasil pengurangan x-x_i, maka selanjutnya dilakukan perhitungan untuk mendapatkan panjang vektor yang digambarkan pada Tabel 3.20. Besar atau panjang vektor misalnya r (x,y,z) dapat ditentukan dengan rumus.

$$|r| = \sqrt{x^2 + y^2 + z^2}$$

Tabel 3.20 Perhitungan panjang vektor untuk K(2,1) dan K(2,2)

||x1-x1||	$\sqrt{\begin{matrix}0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\end{matrix}}$	0
||x1-x2||	$\sqrt{\begin{matrix}0^2+0^2+0^2+0.16903085^2+0^2+0^2+0^2+0^2+0.16903085^2+0^2+\\0^2-0.14002801^2-0.14002801^2-0.14002801^2+0^2+0.16903085^2+0^2\\+0^2+0.16903085^2-0.42008403^2-0.14002801^2+0.16903085^2+0^2+0^2\\+0^2+0^2+0^2-0.14002801^2-0.14002801^2+0^2+\\0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2-0.14002801^2+\\0.16903085^2+0^2-0.14002801^2+0^2-0.14002801^2+0^2+0^2-0.14002801^2\\+0^2+0^2+0^2+0.16903085^2+0^2+0^2-0.14002801^2-0.14002801^2+0^2+0.16903085^2+0^2+(\end{matrix}}$	1.41421
||x2-x1||	$\sqrt{\begin{matrix}0^2+0^2+0^2+0.16903085^2+0^2+0^2+0^2+0^2+0.16903085^2+0^2+\\0^2-0.14002801^2-0.14002801^2-0.14002801^2+0^2+0.16903085^2+0^2\\+0^2+0.16903085^2-0.42008403^2-0.14002801^2+0.16903085^2+0^2+0^2\\+0^2+0^2+0^2-0.14002801^2-0.14002801^2+0^2+\\0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2-0.14002801^2+\\0.16903085^2+0^2-0.14002801^2+0^2-0.14002801^2+0^2+0^2-0.14002801^2\\+0^2+0^2+0^2+0.16903085^2+0^2+0^2-0.14002801^2-0.14002801^2+0^2+0.16903085^2+\end{matrix}}$	1.41421
||x2-x2||	$\sqrt{\begin{matrix}0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\\+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2+0^2\end{matrix}}$	0

Hasil dari panjang vektor pada Tabel 3.20 kemudian dimasukan ke dalam persamaan exp $(-\gamma\|x - x_i\|^2)$ sebagai berikut:

K(1,1) = exp $(-\gamma\|x2 - x1\|^2)$
=exp (-0.5(0^2))
=exp (0)
=1

K(1,2) = exp $(-\gamma\|x2 - x2\|^2)$
=exp (-0.5(1.41421^2))
=exp (-0.999999)
= 0.367879

K(2,1) = exp $(-\gamma\|x2 - x1\|^2)$
=exp (-0.5(1.41421^2))
=exp (-0.999999)
= 0.367879

K(2,2) = exp $(-\gamma\|x2 - x1\|^2)$
=exp (-0.5(0^2))
=exp (0)
=1

Perhitungan untuk nilai exp $(-\gamma\|x - x_i\|^2)$ selanjutnya didapatkan nilai pada Tabel 3.21

Tabel 3.21 Hasil perhitungan nilai K dengan kernel

K(1,1)	K(1,2)	K(2,1)	K(2,2)
1	0.3679	0.3679	1

Setelah dilakukan perhitungan, maka matriks K yang terbentuk adalah sebagai berikut.

$$K = \begin{bmatrix} 1 & 0.3679 \\ 0.3679 & 1 \end{bmatrix}$$

Selanjutnya dilakukan perhitungan $y_i y_j$ sebanyak N data. Hasil dari perhitungan tersebut dapat dilihat pada Tabel 3.22.

Tabel 3.22 Hasil perhitungan $y_i y_j$

$y_1 y_1$	$y_1 y_2$	$y_2 y_1$	$y_2 y_2$
1	-1	-1	1

Sehingga matriks yang terbentuk dari perhitungan $y_i y_j$ hasilnya sebagai berikut:

$$y_i y_j = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$$

Matriks diatas setiap elemenya merupakan hasil exp $(-\gamma \|x - x_i\|^2)$ yang akan berkorelasi dengan $a_i a_j$ dalam persamaan 2.23. Dengan menggunakan kernel K sebagai pengganti dot-product exp $(-\gamma \|x - x_i\|^2)$ dalam persamaan dualitas Lagrange multiplier 2.23, didapatkan:

Maksimalkan:

Ld max = $a_1 + a_2 - \frac{1}{2}({a_1}^2 - 0.73575888 a_1 a_2 + {a_2}^2)$

Syarat 1: $a_1 - a_2 = 0$

Syarat 2: $a_1 , a_2 \geq 0$

Dalam fungsi tujuan, suku kedua sudah dikalikan dengan $y_i y_j$. Persamaan tersebut memenuhi standar *Quadaratic programming* sehingga bisa dibantu penyelesainnya dengan solver komersial untuk *Quadratic Programming* (QP). Dengan bantuan perangkat lunak, didapatkan hasil sebagai berikut.

a_1 = 1.0

a_2 = 1.0

b = -0

Hasil ini menunjukan bahwa semua data *training* adalah *support vector* dikarenakan nilai a > 0. Sementara nilai b didapatkan dari proses pelatihan yang dilakukan. Setelah menemukan semua a dan b, maka model SVM sudah siap digunakan untuk prediksi dengan menggunakan fungsi pemisah berikut [12].

f(x) = w^T. x + b

dimana w^T. x + b = $a^T y \quad K(x_{training}, x_{testing})$+b

Untuk persamaan f(x) adalah sebagai berikut.

f(x) = $a^T y \quad K(x_{training}, x_{testing})$+b

$$= \begin{bmatrix} 1.0 \\ 1.0 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \exp\left(-0.5 \left\| x_{training} - x_{testing} \right\|^2\right) - 0$$

3.1.4.2.3 Analisis Training ECOC

Dilakukan pembangunan model klasifikasi pada data artikel menggunakan SVM untuk 5 kelas yaitu “ekonomi”, ”kesehatan”, ”teknologi”, “pariwisata” dan

“olahraga”. Pendekatan yang dilakukan adalah ECOC. *Codeword* adalah sebuah susunan kode biner untuk merepresentasikan tiap karakter pada suatu data. Panjang *codeword* yang digunakan untuk setiap kelas adalah 3 bit. Seperti ditunjukan pada tabel berikut.

Tabel 3.23 Encoding kelas data artikel

Kelas	Codeword
ekonomi	1 1 -1 -1
kesehatan	1 -1 -1 1
teknologi	-1 1 -1 -1

Parameter kernel yang digunakan dalam contoh ini adalah Gaussian RBF dengan γ (gamma) = 0.5. Selanjutnya dilakukan pembangunan model untuk memisahkan kesehatan dengan {ekonomi, teknologi, pariwisata dan olahraga } dalam hal ini berarti memisahkan Doc2 dengan Doc1, Doc2, Doc4 dan Doc5.

Dalam tahap ini diambil contoh kasus yang telah melalui tahap pembobotan TF-IDF. Dokumen yang digunakan adalah Doc1, Doc2 , Doc3, Doc4 dan Doc5 dengan sudah diberikan label, dimana Doc2 diberikan label +1 sedangkan Doc1 , Doc3, Doc4 dan Doc5 diberikan label -1 pada Tabel 3.11. Gambaran kelas data artikel yang direpresentasikan dalam bentuk Tabel 3.24.

Tabel 3.24 Data Support Vector

	Doc1	Doc2	Doc3	Doc4	Doc5
Y	-1	1	-1	-1	-1

Tahap selanjutnya adalah menghitung nilai x. Nilai x pada tabel akan digunakan untuk perhitungan kernel atau perhitungan dot product. Nilai x didapat dari nilai term pada dokumen, nilai yang digunakan adalah adalah nilai yang bukan bernilai 0. Pada tahap selanjutnya yaitu melakukan kernelisasi menggunakan fungsi kernel RBF yang didefinisikan sebagai K(x,y) = exp $(-\gamma\|x-y\|^2)$, $\gamma > 0$.

Formulasi yang digunakan adalah dualitas Lagrange multiplier(persamaan 2.23) sehingga untuk w harus menggunakan persamaan 2.17 yang sudah dimodifikasi untuk x dengan fungsi kernelnya, menjadi:

W= $\sum_{i=1}^{N} a_i y_i \emptyset x_i$

Selanjutnya harus dilakukan kernelisasi pada set data dari fitur dimensi lama sehingga mendapatkan set data dengan fitur baru dimensi tinggi. Dengan kernel $K(x_i,x_j) = \exp(-\gamma\|x - x_i\|^2)$ dan set data berdimensi Nx1 maka akan didaptakan dimensi baru NxN, dimana N adalah banyaknya data. Untuk hasil perhitungan untuk nilai exp $(-\gamma\|x - x_i\|^2)$ untuk hasil perhitungan nilai K dapat dilihat pada Tabel 3.13.

Tabel 3.25 Hasil perhitungan nilai K dengan kernel

K(1,1)	**K(1,2)**	**K(1,3)**	**K(1,4)**	$\boldsymbol{K(1,5)}$
1	0.367879441	0.367879437	0.36787944	0.367879449
K(2,1)	**K(2,2)**	**K(2,3)**	**K(2,4)**	**K(2,5)**
0.367879441	1	0.367879433	0.367879435	0.367879445
$\boldsymbol{K(3,1)}$	**K(3,2)**	**K(3,3)**	**K(3,4)**	**K(3,5)**
0.367879437	0.367879433	1	0.367879432	0.367879441
K(4,1)	**K(4,2)**	**K(4,3)**	**K(4,4)**	**K(4,5)**
0.36787944	0.367879435	0.367879432	1	0.367879443
K(5,1)	**K(5,2)**	**K(5,3)**	**K(5,4)**	**K(5,5)**
0.367879449	0.367879445	0.367879441	0.367879443	1

Setelah dilakukan perhitungan, maka matriks kernel K yang terbentuk adalah sebagai berikut.

$$K = \begin{bmatrix} \mathrm{K}(1,1) & \mathrm{K}(2,1) & \mathrm{K}(3,1) & \mathrm{K}(4,1) & \mathrm{K}(5,1) \\ \mathrm{K}(1,2) & \mathrm{K}(2,2) & \mathrm{K}(3,2) & \mathrm{K}(4,2) & \mathrm{K}(5,2) \\ \mathrm{K}(1,3) & \mathrm{K}(2,3) & \mathrm{K}(3,3) & \mathrm{K}(4,3) & \mathrm{K}(5,3) \\ \mathrm{K}(1,4) & \mathrm{K}(2,4) & \mathrm{K}(3,4) & \mathrm{K}(4,4) & \mathrm{K}(5,4) \\ \mathrm{K}(1,5) & \mathrm{K}(2,5) & \mathrm{K}(3,5) & \mathrm{K}(4,5) & \mathrm{K}(5,5) \end{bmatrix}$$

$$K = \begin{bmatrix} 1 & 0.367879441 & 0.367879437 & 0.36787944 & 0.367879449 \\ 0.367879441 & 1 & 0.367879433 & 0.367879435 & 0.367879445 \\ 0.367879437 & 0.367879433 & 1 & 0.367879432 & 0.367879441 \\ 0.36787944 & 0.367879435 & 0.367879432 & 1 & 0.367879443 \\ 0.367879449 & 0.367879445 & 0.367879441 & 0.367879443 & 1 \end{bmatrix}$$

Selanjutnya dilakukan perhitungan $y_i y_j$ sebanyak N data. Hasil dari perhitungan tersebut dapat dilihat pada Tabel 3.26

Tabel 3.26 Hasil perhitungan $y_i y_j$

y_1y_1	y_1y_2	y_1y_3	y_1y_4	y_1y_5
1	-1	-1	-1	-1
y_2y_1	y_2y_2	y_2y_3	y_2y_4	y_2y_5
-1	1	1	1	1
y_3y_1	y_3y_2	y_3y_3	y_3y_4	y_3y_5
-1	1	1	1	1
y_4y_1	y_4y_2	y_4y_3	y_4y_4	y_4y_5
-1	1	1	1	1
y_5y_1	y_5y_2	y_5y_3	y_5y_4	y_5y_5
-1	1	1	1	1

Sehingga matriks yang terbentuk dari perhitungan $y_i y_j$ hasilnya sebagai berikut:

$$y_i y_j = \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Matriks K diatas setiap elemenya merupakan hasil exp $(-\gamma\|x - x_i\|^2)$, yang akan berkorelasi dengan $a_i a_j$ dalam persamaan 2.23. Dengan menggunakan kernel K sebagai pengganti dot-product $x_i x_j$ dalam persamaan dualitas Lagrange multiplier 2.23, didapatkan:

Maksimalkan:

Ld max = $a_1 + a_2 + a_3 + a_4 + a_5$ - $\frac{1}{2}$ ($a_1{}^2$ + 0.73575888a_1a_2+ 0.73575888a_1a_3+0.73575888a_1a_4+0.73575888a_1a_5+ $a_2{}^2$ + 0.73575888a_2a_3+0.73575888a_2a_4 +0.73575888a_2a_5+ $a_3{}^2$ +0.73575888a_3a_4+0.73575888a_3a_5+ $a_4{}^2$ +0.73575888a_4a_5+ $a_5{}^2$))

Syarat 1: a_1 - a_2 - a_3 - a_4 - a_5= 0

Syarat 2: a_1 , a_2 , a_3, a_4, a_5 $\geq$ 0

Dalam fungsi tujuan, suku kedua sudah dikalikan dengan $y_i y_j$. Persamaan tersebut memenuhi standar *Quadaratic programming* sehingga bisa dibantu penyelesainnya dengan solver komersial untuk *Quadratic Programming* (QP). Dengan bantuan perangkat lunak, didapatkan hasil sebagai berikut.

a_1 = 0.25

a_2 = 0.25

a_3 = 0.25

a_4 = 0.25

a_5 = 1

b = -0.84196986

Hasil ini menunjukan bahwa semua data *training* adalah *support vector*. Karena nilai a > 0. Sementara nilai b didapatkan dari proses pelatihan yang dilakukan. Setelah menemukan semua a dan b, maka model SVM sudah siap digunakan untuk prediksi dengan menggunakan fungsi pemisah berikut [12`].

f(x) = w^T. x + b

dimana w^T. x + b = $a^T y$ $K(x_{training}, x_{testing})$+b

Untuk persamaan f(x) adalah sebagai berikut.

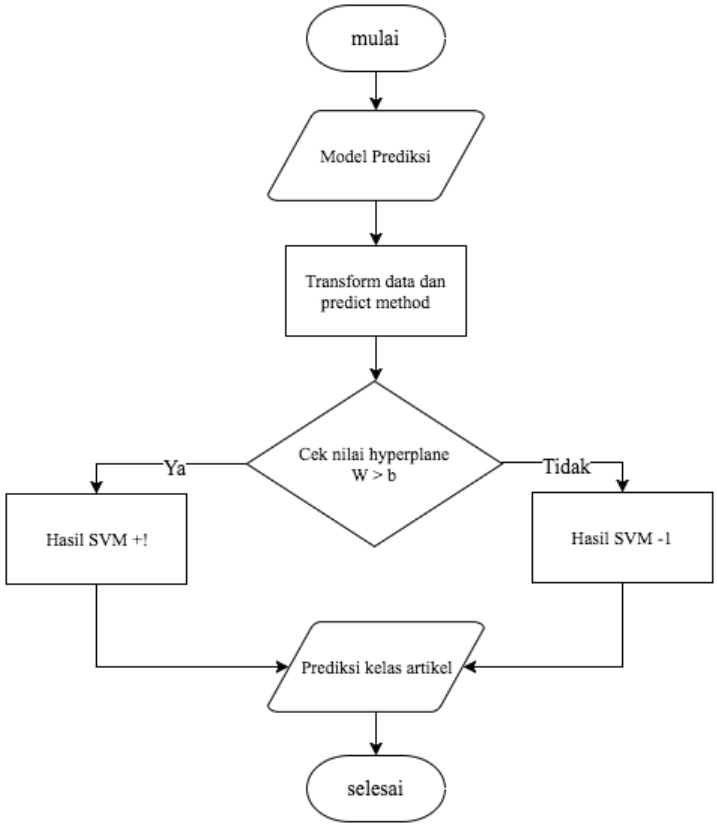
$$f(x) = a^T y \quad K(x_{training}, x_{testing}) + b$$

$$= \begin{bmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \\ 1.0 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \end{bmatrix} \exp\left(-0.5 \left\| x_{training}\ x_{testing} \right\|^2\right) +$$

(-0.84196986)

3.1.4.3 Analisis Proses SVM Testing

Setelah mendapatkan model prediksi, selanjutnya dapat menentukan data artikel masuk ke dalam kelas positif atau negatif dengan nilai w dan *hyperplane* dari data *training*. Jika nilai hasil uji lebih besar dari nilai *hyperplane* maka abstrak tersebut masuk dalam kelas positif, jika lebih kecil dari nilai *hyperplane* maka abstrak tersebut masuk dalam kelas negatif. Berikut alur proses dapat dilihat pada Gambar 3.10.



Gambar 3.13 *Flowchart* SVM *Testing*

Sebagai contoh, data satu uji dilakukan pada sebuah Artikel *Testing* yang kemudian akan diolah melalui *Preprocessing* dan Pembobotan TF-IDF.

Contoh data artikel yang akan dijadikan data latih, yaitu:

" - Perusahaan otomotif Volkswagen (VW) menyatakan bakal menaikkan gaji kepada 120.000 orang karyawannya di Jerman sebesar 5 persen. Kenaikan gaji ini akan dilakukan dalam kurun 1,5 tahun mendatang. Padahal, saat ini VW tengah mengalami permasalahan terkait skandal kecurangan emisi. Namun demikian, skandal ini tampaknya tak menghentikan keputusan VW untuk menaikkan gaji karyawan."

Term yang didapatkan dari artikel diatas dapat dilihat pada Tabel 3.27

Tabel 3.27 Term data uji

Term **Data uji**	mengalami , global, orang, metall, emisi, vw, perusahaan, jerman, ig, group, otomotif, permasalahan, kerja, dikonsultasikan , jumat, keputusan, persen, pekerja, menaikkan, terkait, volkswagen, kurun, tenaga, skandal, kecurangan, berlaku, karyawannya, karyawan, gaji, kenaikan, menghentikan, serikat

Data *Term* data uji akan digunakan untuk mencari nilai $x_{testing}$.

3.1.4.3.1 Analisis Multiclass *One vs All*

Proses klasifikasi menggunakan metode SVM membagi 2 kelas artikel menjadi kelas positif dan negatif. Ada 2 proses yang dilakukan dalam mengklasifikasi , yaitu proses *training* dan *testing.* Proses *training* dilakukan untuk menghasilkan model prediksi yang akan digunakan dalam menentukan klasifikasi artikel kedalam kelas positif atau negatif.

Seperti yang telah dijelaskan sebelumnya, SVM sebenarnya digunakan untuk klasifikasi biner (dua kelas), dalam kasus multiclass, SVM yang digunakan adalah dengan mengkombinasikan beberapa SVM biner tersebut.

Metode “One-against-all” salah satunya. Dibangun sejumlah *k* SVM biner, dengan *k* adalah jumlah kelas. Untuk setiap model klasifikasi ke-i dilatih dengan menggunakan keseluaruhan data, untuk mencari solusi permasalahan Contohnya, untuk persoalan klasifikasi dengan 5 buah jumlah kelas, digunakan 5 buah SVM biner pada Tabel (2-4) dan penggunaannya pada pengklasifikasian data baru. Kemudian kelas dari suatu data/obyek baru x ditentukan berdasarkan nilai terbesar dari fungsi pemisah.

DocTest yang diuji akan dihitung berdasarkan data artikel *training* dalam kasus ini adalah Doc1, Doc2 dan Doc3 yang ada pada Tabel 3.11. Berikut gambaran sebagian vektor artikel yang digambarkan dalam bentuk tabel 3.28. Untuk hasil selengkapnya dapat dilihat pada Lampiran F

Tabel 3. 28 Vektor Artikel

	Doc1	Doc2	Doc3	Doc4	Doc5	DocTest
term1	0	0	0	0.09950372	0	0
term2	0	0	0	0.09950372	0	0
term3	0	0	0	0.09950372	0	0
term4	0.16903085	0	0	0	0	0
term5	0	0	0	0.19900744	0	0
term6	0	0	0	0.09950372	0	0
term7	0	0	0	0.09950372	0	0
term8	0	0	0	0.39801488	0	0
term9	0.16903085	0	0	0	0	0
term10	0	0	0	0.09950372	0	0
term11	0	0	0	0	0.15075567	0
term12	0	0.14002801	0	0	0	0
term13	0	0.14002801	0	0	0	0
term14	0	0.14002801	0	0	0	0
term15	0	0	0	0.09950372	0	0
term16	0.16903085	0	0	0	0	0
term17	0	0	0	0.29851116	0	0
term18	0	0	0	0.09950372	0	0
term19	0.16903085	0	0	0	0	0
term20	0	0.42008403	0	0	0	0
term21	0	0.14002801	0	0	0	0
term22	0.16903085	0	0	0	0	0

term23	0	0	0	0.09950372	0	0
term24	0	0	0	0.09950372	0	0
term25	0	0	0	0	0.15075567	0
term26	0	0	0	0	0.15075567	0
term27	0	0	0	0	0.15075567	0
term28	0	0.14002801	0	0	0	0
term29	0	0.14002801	0	0	0	0
term30	0	0	0	0	0.30151134	0
term31	0	0	0	0	0.15075567	0
term32	0	0	0	0	0.15075567	0
term33	0	0	0	0.09950372	0	0
term34	0	0	0	0	0.15075567	0
term35	0	0	0	0	0.15075567	0
term36	0	0	0.17677	0	0	0
term37	0	0	0.17677	0	0	0
term38	0	0	0.17677	0	0	0
term39	0	0	0	0	0.15075567	0
term40	0	0.14002801	0	0	0	0
term41	0.16903085	0	0	0	0	0
term42	0	0	0	0.29851116	0	0
term43	0	0.14002801	0	0	0	0
term44	0	0	0.17677	0	0	0
term45	0	0.14002801	0	0	0	0
term46	0	0	0	0.09950372	0	0.15075567
term47	0	0	0.17677	0	0	0
term48	0	0.14002801	0	0	0	0
term49	0	0	0.17677	0	0	0
term50	0	0	0	0	0.15075567	0
term51	0	0	0.17677	0	0	0
term52	0.16903085	0	0	0	0	0
term53	0	0	0	0.09950372	0	0
term54	0	0	0	0	0.15075567	0
term55	0	0.14002801	0	0	0	0
term56	0	0.14002801	0	0	0	0.15075567
term57	0	0	0	0	0.15075567	0
term58	0.16903085	0	0	0	0	0.15075567
term59	0	0	0	0	0.15075567	0
term60	0	0	0.35355	0	0	0

Pada Tabel 3.29 menunjukan semua fungsi pemisah dari 5 kelas yang berarti ada 5 fungsi pemisah yang memisahkan satu kelas dengan 4 kelas lainnya. $K(x_{training}, x_{testing})$ merupakan fungsi kernel RBF exp $(-\gamma\|x-y\|^2)$, $\gamma > 0$ dengan nilai a dan b diperoleh dari proses SVM training dan y merupakan hasil perhitungan dari $y_i y_j$.

Tabel 3.29 Model Klasifikator

y_i = -1	y_i = 1	Hipotesis
ekonomi	Bukan ekonomi	$f(x)=a^T y \quad K(x_{training}, x_{testing})$+b $= \begin{bmatrix} 1.0 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \end{bmatrix}$ $\exp\left(-0.5\|x_{training} - x_{testing}\|^2\right)$ + 0.84196986
kesehatan	Bukan kesehatan	$f(x)=a^T y \quad K(x_{training}, x_{testing})$+b $= \begin{bmatrix} 1.0 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \end{bmatrix}$ $\exp\left(-0.5\|x_{training} - x_{testing}\|^2\right)$ + 0.84196986
teknologi	Bukan teknologi	$f(x)=a^T y \quad K(x_{training}, x_{testing})$+b $= \begin{bmatrix} 1.0 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \end{bmatrix}$ $\exp\left(-0.5\|x_{training} - x_{testing}\|^2\right)$ + 0.84196986
pariwisata	Bukan pariwisata	$f(x)=a^T y \quad K(x_{training}, x_{testing})$+b

		$= \begin{bmatrix} 1.0 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \end{bmatrix}$ $\exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right)+$ 0.84196986
olahraga	Bukan Olahraga	$f(x)=a^T y^T K(x_{training}, x_{testing})$+b $= \begin{bmatrix} 1.0 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \end{bmatrix}$ $\exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right) +$ 0.84196986

$x_{training}$ merupakan data vektor training, misalnya data Doc1 pada Tabel 3.28 dan $x_{testing}$ merupakan data vektor testing, misalnya data DocTest pada Tabel 3.28. Selanjutnya dilakukan perhitungan fungsi pemisah dengan melakukan kernelisasi menggunakan fungsi kernel RBF yang didefinisikan sebagai K(x,y) = exp $(-\gamma\|x-y\|^2), \gamma > 0$.

Tabel 3.30 Hasil perhitungan x - x_i untuk K($x_{training}, x_{testing}$) untuk fungsi pemisah ekonomi

$x_{training} - x_{training}$	$x_{training} - x_{testing}$
0	0.16903085
0	0.16903085
0	0
0	0
0	0
0	0.16903085
0	0.16903085
0	0
0	0
0	0.16903085
0	0

0	0
0	0
0	0
0	0
0	0
0	0.16903085
0	0
0	0
0	0
0	0
0	0
0	0
0	0
0	0.16903085
0	0
0	0
0	0.16903085
0	0
0	0.16903085
0	0.16903085
0	0.16903085
0	0.16903085
0	0.16903085
0	0
0	0
0	0
0	0
0	0
0	0
0	0.16903085
0	0
0	0.16903085
0	0
0	0
0	-0.15075567
0	0

0	0.16903085
0	0
0	0
0	0
0	0
0	0.16903085
0	0
0	0.16903085
0	0.01827518
0	0.16903085
0	-0.15075567
0	0.16903085
0	0
0	0.50709255
0	0
0	0
0	0.3380617
0	0
0	0.16903085
0	0
0	0
0	0
0	0
0	0
0	0

Setelah didapatkan hasil pengurangan x-x_i, maka selanjutnya dilakukan perhitungan untuk mendapatkan panjang vektor yang digambarkan pada Tabel 3.31.

Tabel 3.31 Perhitungan panjang vektor untuk K($x_{training}, x_{testing}$)

$\|\|x_{training} - x_{testing}\|\|$	$\sqrt{\begin{array}{l}0.16903085^2 + 0.16903085^2 + 0^2 + 0^2 + 0^2 \\ +0.16903085^2 + 0.16903085^2 + 0^2 + 0^2 + \\ 0.16903085^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + \\ 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + \\ 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + \\ 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + \\ 0.16903085^2 + 0.0285^2 + 0.1142^2 + 0^2\end{array}}$	=2.02

K($x_{training}, x_{testing}$) = exp $(-\gamma\|x2 - x1\|^2)$

=exp (-0.5(2.02^2))

=exp (-2.0402)

= 0.130002708

Nilai prediksi dari setiap fungsi pemisah diberikan dalam matriks berikut. Setiap kolom dari matriks mewakili nilai prediksi dari setiap fungsi pemisah.

y = $[0.13000278 \quad 1.3010 \quad -9.076 \quad 0.03679 \quad -0.7588]$

Berdasarkan nilai maksimum y, maka hasil dari klasifikasi multiclass adalah bahwa data testing merupakan bagian dari kelas kesehatan.

3.1.4.3.2 Analisis Multiclass *One vs One*

Metode "One-against-one" dibangun sejumlah $\frac{k(k-1)}{2}$ buah SVM biner, dengan k adalah jumlah kelas. Untuk setiap model klasifikasi ke-i dilatih dengan menggunakan keseluaruhan data, untuk mencari solusi permasalahan Contohnya, untuk persoalan klasifikasi dengan 5 buah jumlah kelas, digunakan 10 buah SVM biner pada Tabel 3.32 dan penggunaannya pada pengklasifikasian data baru. $K(x_{training}, x_{testing})$ merupakan fungsi kernel RBF exp $(-\gamma\|x - y\|^2), \gamma >$ 0 dengan nilai a dan b diperoleh dari proses SVM training dan y merupakan hasil perhitungan dari $y_i y_j$.

Tabel 3.32 Model Klasifikator

Pasangan kelas biner	Hipotesis
ekonomi = +1 kesehatan = -1	$f(x)=a^T y \quad K(x_{training}, x_{testing})+b$ $= \begin{bmatrix}1.0\\1.0\end{bmatrix}\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} \exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right)- 0$
ekonomi = 1+ teknologi = 1-	$f(x)=a^T y \quad K(x_{training}, x_{testing})+b$ $= \begin{bmatrix}1.0\\1.0\end{bmatrix}\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} \exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right)- 0$
ekonomi = 1+ pariwisata = 1-	$f(x)=a^T y \quad K(x_{training}, x_{testing})+b$ $= \begin{bmatrix}1.0\\1.0\end{bmatrix}\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} \exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right)- 0$
ekonomi = 1+ olahraga = 1-	$f(x)=a^T y \quad K(x_{training}, x_{testing})+b$ $= \begin{bmatrix}1.0\\1.0\end{bmatrix}\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} \exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right)- 0$
kesehatan = 1+ teknologi = 1-	$f(x)=a^T y \quad K(x_{training}, x_{testing})+b$ $= \begin{bmatrix}1.0\\1.0\end{bmatrix}\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} \exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right)- 0$
kesehatan = 1+ pariwisata = 1-	$f(x)=a^T y \quad K(x_{training}, x_{testing})+b$ $= \begin{bmatrix}1.0\\1.0\end{bmatrix}\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} \exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right)- 0$
kesehatan = 1+ olahraga = 1-	$f(x)=a^T y \quad K(x_{training}, x_{testing})+b$ $= \begin{bmatrix}1.0\\1.0\end{bmatrix}\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} \exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right)- 0$
teknologi = 1+ pariwisata = 1-	$f(x)=a^T y \quad K(x_{training}, x_{testing})+b$ $= \begin{bmatrix}1.0\\1.0\end{bmatrix}\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} \exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right)- 0$

teknologi = 1+ olahraga = 1-	$f(x)=a^T y \quad K(x_{training}, x_{testing})$+b $= \begin{bmatrix}1.0\\1.0\end{bmatrix}\begin{bmatrix}1 & -1\\-1 & 1\end{bmatrix} \exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right)$- 0

$x_{training}$ merupakan data vektor training, misalnya data Doc1 pada Tabel 3.28 dan $x_{testing}$ merupakan data vektor testing, misalnya data DocTest pada Tabel 3.28. Selanjutnya dilakukan perhitungan fungsi pemisah dengan melakukan kernelisasi menggunakan fungsi kernel RBF yang didefinisikan sebagai K(x,y) = exp $(-\gamma\|x-y\|^2), \gamma > 0$.

Tabel 3.33 Hasil perhitungan x - x_i untuk K($x_{training}, x_{testing}$)

$x_{training} - x_{testing}$
0.16903085
0.16903085
0
0
0
0.16903085
0.16903085
0
0
0.16903085
0
0
0
0
0
0
0.16903085
0
0
0
0
0
0

0
0.16903085
0
0
0.16903085
0
0.16903085
0.16903085
0.16903085
0.16903085
0.16903085
0
0
0
0
0
0
0.16903085
0
0.16903085
0
0
-0.15075567
0
0.16903085
0
0
0
0
0.16903085
0
0.16903085
0.01827518
0.16903085
-0.15075567
0.16903085

0
0.50709255
0
0
0.3380617
0
0.16903085
0
0
0
0
0
0

Setelah didapatkan hasil pengurangan x-x_i, maka selanjutnya dilakukan perhitungan untuk mendapatkan panjang vektor yang digambarkan pada Tabel 3.34.

Tabel 3.34 Perhitungan panjang vektor untuk K($x_{training}, x_{testing}$)

$\|\|x_{training} - x_{testing}\|\|$	$\sqrt{\begin{array}{l}0.16903085^2 + 0.16903085^2 + 0^2 + 0^2 + 0^2 \\ +0.16903085^2 + 0.16903085^2 + 0^2 + 0^2 + \\ 0.16903085^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + \\ 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + \\ 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + \\ 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + \\ 0.16903085^2 + 0.0285^2 + 0.1142^2 + 0^2\end{array}}$	=2.02

Kemudian prediksikan x berada dalam kelas mana, didasarkan pada nilai voting yang paling tinggi [22]. Dalam kasus dimana ada dua kelas dengan nilai voting yang sama, pilih yang indeksnya paling kecil[22].

Hasil dari klasifikasi multiclass adalah bahwa data testing merupakan bagian dari kelas olahraga.

3.1.4.3.3 Analisis Multiclass *Error Corecting Output Code*

Untuk masalah multikelas, setiap kelas y_i direpresentasikan oleh string bit dengan panjang n yang disebut codewrord. Selanjutnya harus dibentuk n

klasifikator biner untuk memprediksi setiap bit sesuai string codeword. Kelas vector uji hasil prediksi diberikan oleh codeword dengan jarak hamming yang paling dekat terhadap codeword yang dihasilkan oleh klasifikator biner. Codeword adalah *information digit* yang merupakan elemen dari kode.

Kelas	**Codeword**
ekonomi	1 1 -1 -1
kesehatan	1 -1 -1 1
teknologi	-1 1 -1 -1

Tabel 3.35 menunjukan semua fungsi pemisah dari 5 kelas yang berarti ada 5 fungsi pemisah. $K(x_{training}, x_{testing})$ merupakan fungsi kernel RBF exp $(-\gamma\|x-y\|^2), \gamma > 0$ dengan nilai *a* dan b diperoleh dari proses SVM training dan *y* merupakan hasil perhitungan dari $y_i y_j$.

Tabel 3.35 Model Klasifikator

y_i = -1	y_i = 1	Hipotesis
ekonomi	Bukan ekonomi	*f(x)*=$a^T y \quad K(x_{training}, x_{testing})$+b $= \begin{bmatrix} 1.0 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \end{bmatrix}$ $\exp\left(-0.5\|x_{training} - x_{testing}\|^2\right)$ + 0.84196986
kesehatan	Bukan kesehatan	*f(x)*=$a^T y \quad K(x_{training}, x_{testing})$+b

		$= \begin{bmatrix} 1.0 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \end{bmatrix}$ $\exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right)$ + 0.84196986
teknologi	Bukan teknologi	$f(x)=a^T y \quad K(x_{training}, x_{testing})$+b $= \begin{bmatrix} 1.0 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \end{bmatrix}$ $\exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right)$ + 0.84196986
pariwisata	Bukan pariwisata	$f(x)=a^T y \quad K(x_{training}, x_{testing})$+b $= \begin{bmatrix} 1.0 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \end{bmatrix}$ $\exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right)$ + 0.84196986
olahraga	Bukan Olahraga	$f(x)=a^T y \quad K(x_{training}, x_{testing})$+b $= \begin{bmatrix} 1.0 \\ 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 & -1 & -1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \\ -1 & 1 & 1 & 1 & 1 \end{bmatrix}$ $\exp\left(-0.5\left\|x_{training} - x_{testing}\right\|^2\right)$ + 0.84196986

$x_{training}$ merupakan data vektor training, misalnya data Doc1 pada Tabel 3.28 dan $x_{testing}$ merupakan data vektor testing, misalnya data DocTest pada Tabel 3.28. Selanjutnya dilakukan perhitungan fungsi pemisah dengan melakukan kernelisasi

menggunakan fungsi kernel RBF yang didefinisikan sebagai K(x,y) = exp $(-\gamma\|x-y\|^2), \gamma > 0$.

Tabel 3.36 Hasil perhitungan x - x_i untuk K($x_{training}, x_{testing}$)

$x_{training} - x_{testing}$
0.16903085
0.16903085
0
0
0
0.16903085
0.16903085
0
0
0.16903085
0
0
0
0
0
0
0.16903085
0
0
0
0
0
0
0
0.16903085
0
0
0.16903085
0
0.16903085
0.16903085

0.16903085
0.16903085
0.16903085
0
0
0
0
0
0
0.16903085
0
0.16903085
0
0
-0.15075567
0
0.16903085
0
0
0
0
0.16903085
0
0.16903085
0.01827518
0.16903085
-0.15075567
0.16903085
0
0.50709255
0
0
0.3380617
0
0.16903085
0

0
0
0
0
0

Setelah didapatkan hasil pengurangan x-x_i, maka selanjutnya dilakukan perhitungan untuk mendapatkan panjang vektor yang digambarkan pada Tabel 3.37.

Tabel 3.37 Perhitungan panjang vektor untuk K($x_{training}, x_{testing}$)

$\|\|x_{training} - x_{testing}\|\|$	$\sqrt{\begin{array}{l}0.16903085^2 + 0.16903085^2 + 0^2 + 0^2 + 0^2 \\ +0.16903085^2 + 0.16903085^2 + 0^2 + 0^2 + \\ 0.16903085^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + 0^2 + \\ 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + \\ 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + \\ 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + 0.16903085^2 + \\ 0.16903085^2 + 0.0285^2 + 0.1142^2 + 0^2\end{array}}$	=2.02

K($x_{training}, x_{testing}$) = exp $(-\gamma\|x2 - x1\|^2)$

=exp (-0.5(2.02^2))

=exp (-2.0402)

= 0.130002708

Selanjutnya lakukan perhitungan jarak hamming di antara hasil prediksi dengan codeword dari 5 kelas. Hasil dari klasifikasi multiclass adalah bahwa data testing merupakan bagian dari kelas ekonomi.

3.1.5 Analisis kebutuhan non fungsionalitas

Analisis kebutuhan non fungsional adalah langkah dimana seorang pembangun perangkat lunak menganalisis sumber daya yang akan digunakan dan menggunakan perangkat lunak yang dibangun. Perangkat keras dan perangkat lunak yang dimiliki harus sesuai dengan kebutuhan, sehingga dapat ditentukan kompabilitas aplikasi yang dibangun terhadap sumber daya yang ada.

3.1.5.1 Analisis Kebutuhan Perangkat Lunak

Adapun perangkat lunak yang digunakan pada penelitian implementasi klasifikasi teks pada artikel berita berbahasa indonesia menggunakan metode *Multiclass SVM*.

1. Pycharm
2. Robomongo
3. Python 2.7
4. Nltk data
5. Scikit learn
6. Omnigraffle
7. Common Gateaway Interface (CGI)
8. XAMPP

3.1.5.2 Analisis kebutuhan perangkat keras

Adapun perangkat keras yang digunakan pada penelitian klasifikasi teks pada artikel berita berbahasa indonesia menggunakan metode *Multiclass SVM* :

1. Processor P4
2. Ram 512MB
3. Harddisk 40 GB
4. Monitor 13.3 inch
5. Sistem Operasi Windows 7 Profesional 32bit

3.1.5.3 Analisis Pengguna

Dalam penelitian ini hanya akan melibatkan satu pengguna, yaitu penguji. Dengan karakteristik sebagai berikut:

Tabel 3.38 Analisis Pengguna

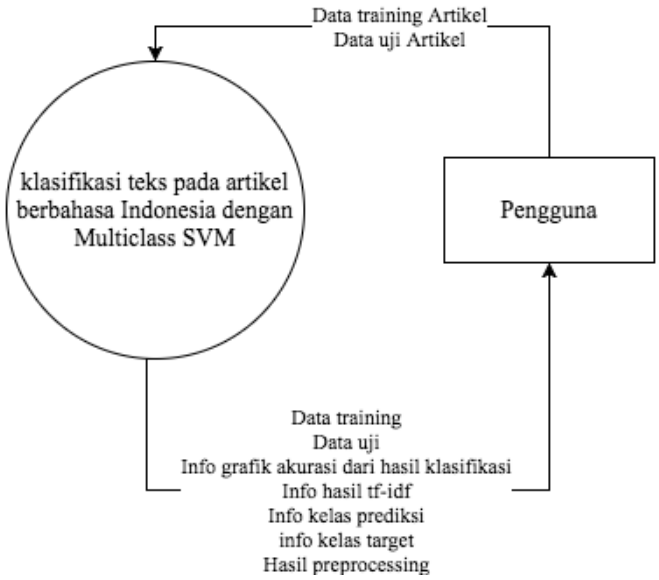
Pengguna	Penguji
Hak Akses	a. Melakukan uji coba data training artikel b. Melakukan uji coba data uji artikel

3.1.6 Analisis Kebutuhan Fungisional

Tujuan dari analisis ini adalah untuk menganalisis proses yang akan diimplementasikan pada penelitian ini menggunakan Diagram Konteks, Pembuatan DFD, dan Spesifikasi Proses.

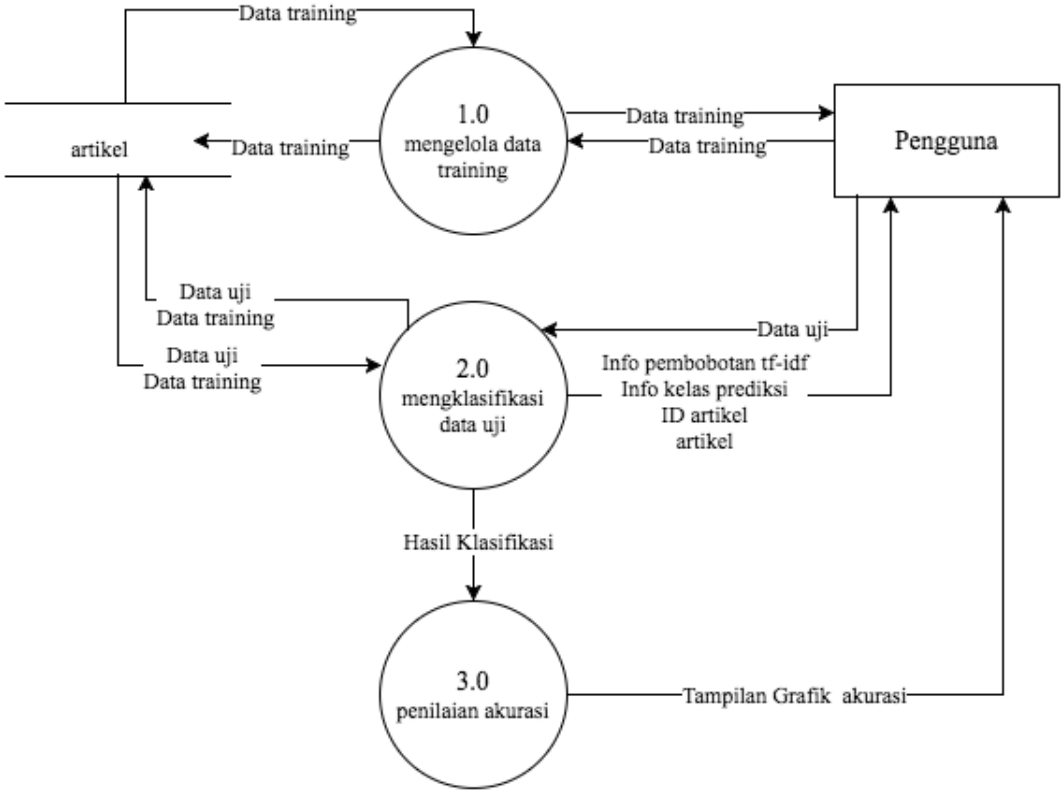
3.1.6.1 Diagram Konteks

Diagram Konteks menggambarkan ruang lingkup sistem. Diagram konteks yang terdapat dalam sistem digambarkan pada Gambar 3.14 Diagram Konteks sebagai berikut:



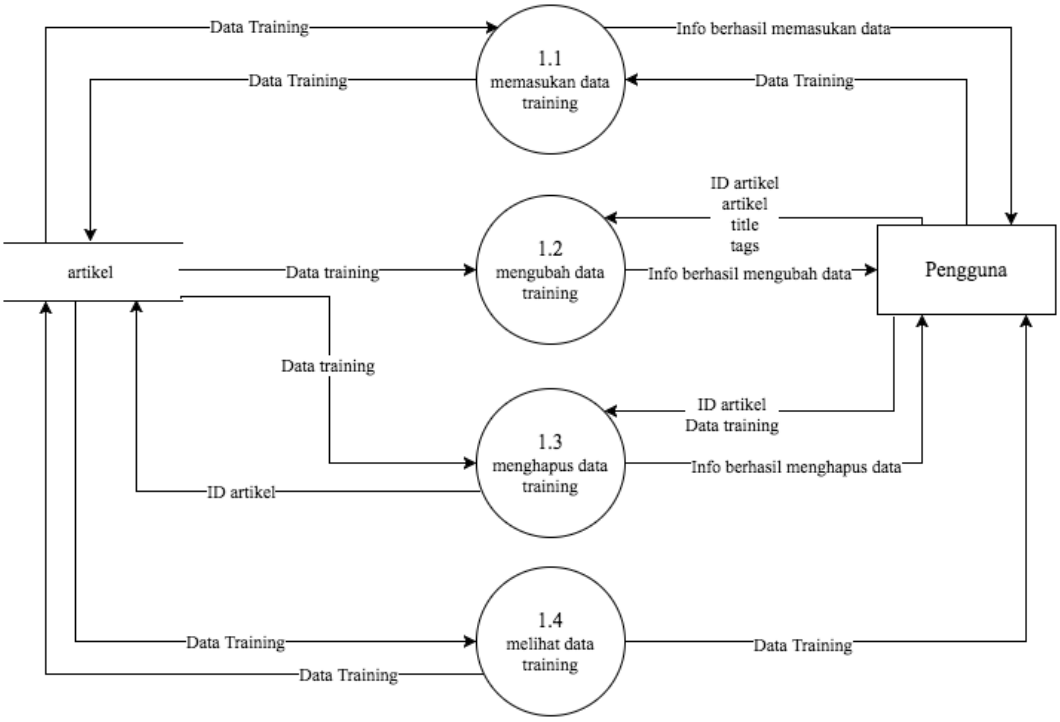
Gambar 3.14 Diagram Konteks

3.1.6.2 Data Flow Diagram (DFD) Level 1



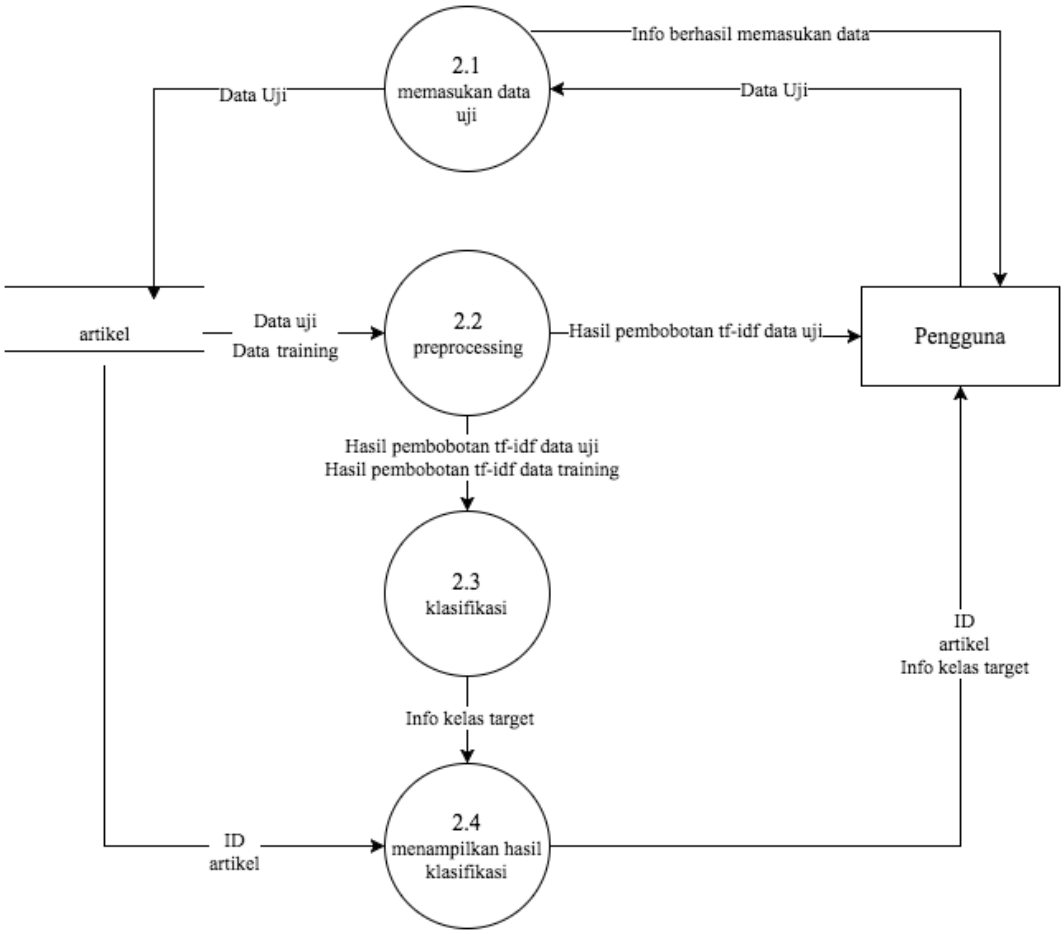
Gambar 3.15 Data Flow Diagram (DFD) Level 1

3.1.6.3 Data Flow Diagram (DFD) Level 2 Mengelola Data Training



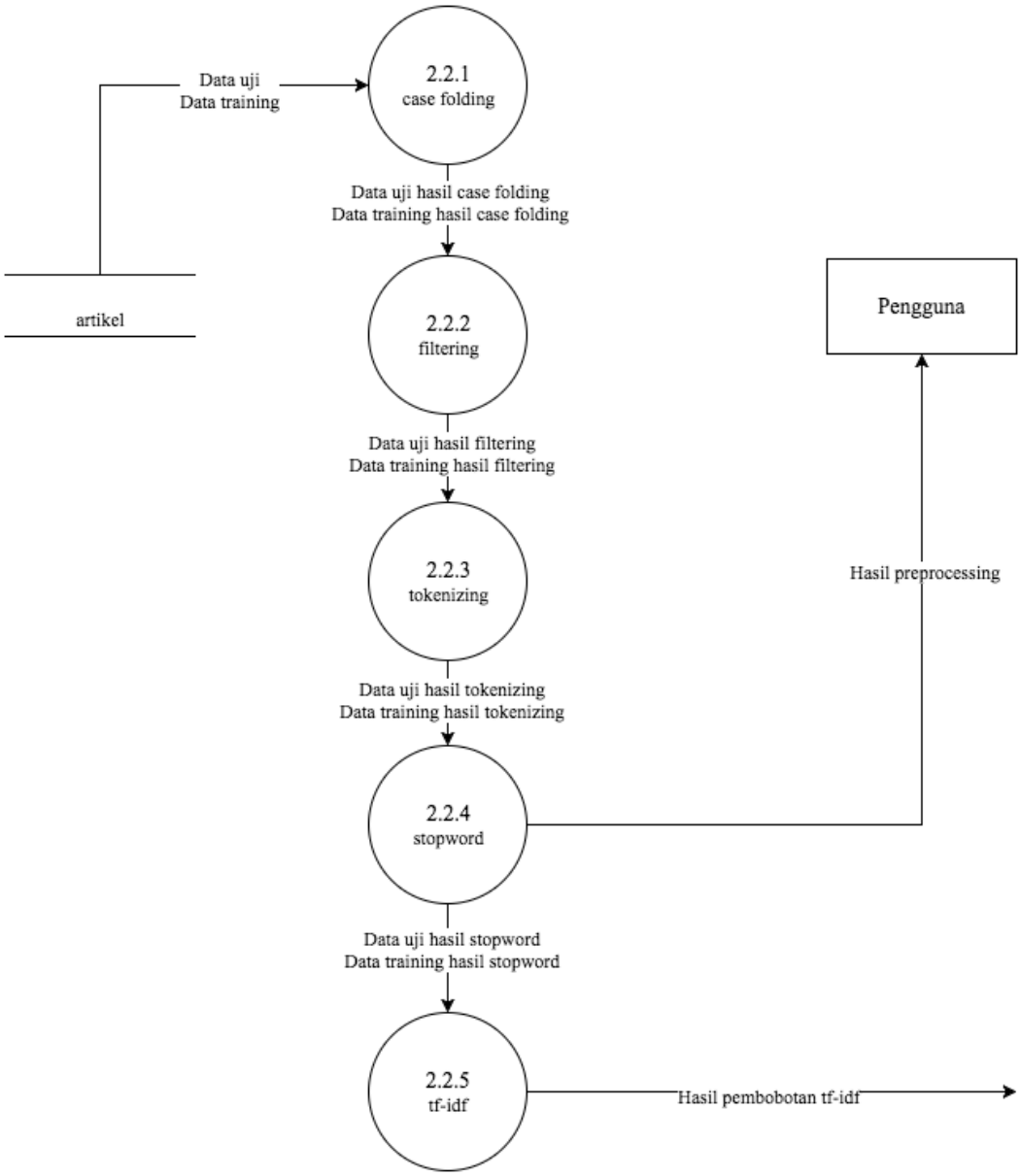
Gambar 3.16 Level 2 Mengelola Data Training

3.1.6.4 Data Flow Diagram (DFD) Level 2 Mengklasifikasi Data Uji



Gambar 3.17 Level 2 Mengklasifikasi Data Uji

3.1.6.5 Data Flow Diagram (DFD) Level 3 Preprocessing



Gambar 3.18 Level 3 Preprocessing

3.1.6 Spesifikasi Proses

Berikut ini ada spesifikasi proses untuk penelitian Implementasi Klasifikasi Teks Pada Artikel Berbahasa Indonesia dengan Multiclass SVM:

Tabel 3.39 Spesifikasi Proses

No	Proses	Keterangan
1	No. Proses	1.0
	Nama	Mengelola Data Training
	Source (sumber)	Pengguna
	Input	Data Training
	Output	Data Training
	Logika Proses	1. Pengguna memasukkan data training 2. Sistem menyimpan data training yang telah di inputkan . 3. Hasil output berupa data training
No	**Proses**	**Keterangan**
2	No. Proses	2.0
	Nama	Mengklasifikasi data uji
	Source (sumber)	Pengguna
	Input	Data Uji dan Data Training
	Output	Data Uji (ID, artikel), hasil pembobotan TF-IDF dan info kelas
	Logika Proses	1. Pengguna memasukan data uji 2. Sistem menyimpan data training yang telah di inputkan. 3. Selanjutnya dilakukan proses preprocessing. 4. Sistem melakukan proses klasifikasi. 5. Hasil output adalah data ID, artikel, kelas data uji dan hasil pembobotan TF-IDF.

No	Proses	Keterangan
3	No. Proses	3.0
	Nama	Penilaian Akurasi
	Source (sumber)	Pengguna
	Input	Hasil klasifikasi
	Output	Tampilan Grafik
	Logika Proses	1. Sistem akan mengambil hasil dari proses klasifikasi. 2. Hasil output berupa grafik perbandingan akurasi, *precission*, *recall*, diantara pendekata metode *Multiclass* SVM (*one vs one, one vs all, error correcting output code*)
No	**Proses**	**Keterangan**
4	No. Proses	1.1
	Nama	Memasukan data training
	Source (sumber)	Pengguna
	Input	Data training
	Output	Info berhasil memasukan data
	Logika Proses	1. Pengguna memasukan data training. 2. Sistem akan memberikan output berupa pesan apabila data berhasil dimasukan dan pesan gagal jika data uji gagal dimasukan ke dalam database.
No	**Proses**	**Keterangan**
5	No. Proses	1.2

	Nama	Mengubah data training
	Source (sumber)	Pengguna
	Input	ID, artikel, title, tags
	Output	Info berhasil mengubah data
	Logika Proses	1. Pengguna mengubah data training yang sudah dipilih. 2. Sistem akan memberikan output berupa pesan apabila data berhasil diubah dan pesan gagal jika data uji gagal diubah.
No	**Proses**	**Keterangan**
6	No. Proses	1.3
	Nama	Menghapus data training
	Source (sumber)	Pengguna
	Input	ID data training, data training
	Output	Info berhasil menghapus data
	Logika Proses	1. Pengguna akan memilih data yang akan dihapus atau menghapus semua data training. 2. Sistem akan memberikan output berupa pesan apabila data berhasil dihapus dan pesan gagal jika data uji gagal dihapus.
No	**Proses**	**Keterangan**
7	No. Proses	1.4
	Nama	Melihat data training
	Source (sumber)	Pengguna

	Input	Data training
	Output	Data training
	Logika Proses	1. Sistem mengambil data training dari database. 2. Hasil output berupa data training.
No	**Proses**	**Keterangan**
8	No. Proses	2.1
	Nama	Memasukan data uji
	Source (sumber)	Pengguna
	Input	Data Uji
	Output	Info berhasil memasukan data
	Logika Proses	1. Pengguna memasukan data uji. 2. Sistem akan memberikan output berupa pesan apabila data berhasil dimasukan dan pesan gagal jika data uji gagal dimasukan ke dalam database.
No	**Proses**	**Keterangan**
9	No. Proses	2.2
	Nama	Preprocessing
	Source (sumber)	Pengguna
	Input	Data training, data uji
	Output	Hasil pembobotan tf-idf data uji dan data training
	Logika Proses	1. Sistem mengambil data uji dan data training. 2. Sistem akan memberikan output berupa hasil pembobotan tf-idf data uji dan data training.

No	Proses	Keterangan
10	No. Proses	2.3
	Nama	Klasifikasi
	Source (sumber)	Pengguna
	Input	Hasil pembobotan tf-idf data uji dan data training
	Output	Info kelas target
	Logika Proses	1. Sistem akan menerima hasil pembobotan tf-idf data uji dan data training dari proses preprocessing. 2. Hasil output berupa info kelas target.
No	**Proses**	**Keterangan**
11	No. Proses	2.4
	Nama	Menampilkan hasil klasifikasi
	Source (sumber)	Pengguna
	Input	Info kelas target, ID, artikel
	Output	ID, artikel, hasil pembobotan tf-idf data uji, info kelas target
	Logika Proses	1. Sistem akan mengambil info kelas target dari proses klasifikasi. 2. Hasil output berupa ID, artikel, hasil pembobotan tf-idf data uji, info kelas target

3.1.7 Kamus Data

Berikut ini adalah kamus data untuk implementasi klasifikasi teks dengan multiclass SVM pada artikel berbahasa Indonesia:

Tabel 3.40 Kamus Data

1.	Nama	:	Data training
	Digunakan di	:	Proses mengelola data training (input / output)
	Deskripsi	:	Data training terdiri dari id, artikel, tags
2.	Nama	:	Data Uji
	Digunakan di	:	Proses mengelola data uji (input / output)
	Deskripsi	:	Data uji terdiri dari id, artikel
3.	Nama	:	Hasil pembobotan TF-IDF
	Digunakan di	:	Proses mengelola ujian (output)
	Deskripsi	:	Hasil pembobotan TF-IDF
4.	Nama	:	Grafik klasifikasi
	Digunakan di	:	Proses grafik klasifikasi
	Deskripsi	:	Proses grafik klasifikasi terdiri dari data uji (id, artikel)

3.2 Perancangan Sistem

Dari hasil analisis yang telah dilakukan, maka dibuatlah perancangan sistem. Perancangan sistem adalah rancangan atau susunan dari analisis yang telah dilakukan. Perencanaan sistem yang dibangun terdiri dari basis data dan perancangan antarmuka.

3.2.1 Data Model

Data model adalah cara formal menggambarkan data yang digunakan dan diciptakan dalam sistem. Pemodelan data NoSQL sering membutuhkan pemahaman yang lebih, dalam struktur data dan algoritma dibandingkan dengan pemodelan database relasional.

```
{
        _id    : <ObjectId1>,
        item_title    : [“Dilanda Skandal, VW Tetap Beri Kenaikan Gaji 5 Persen untuk 120.000 Karyawan”],
```

item_artikel : [“Perusahaan otomotif Volkswagen (VW) menyatakan bakal menaikkan gaji kepada 120.000 orang karyawannya di Jerman sebesar 5 persen. Kenaikan gaji ini akan dilakukan dalam kurun 1,5 tahun mendatang.",

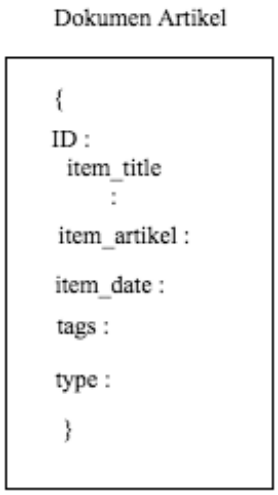
"Padahal, saat ini VW tengah mengalami permasalahan terkait skandal kecurangan emisi. Namun demikian, skandal ini tampaknya tak menghentikan keputusan VW untuk menaikkan gaji karyawan.”]

item_date : [“Juni, 2016”]

item_tags : “ekonomi”,

type : “testing”

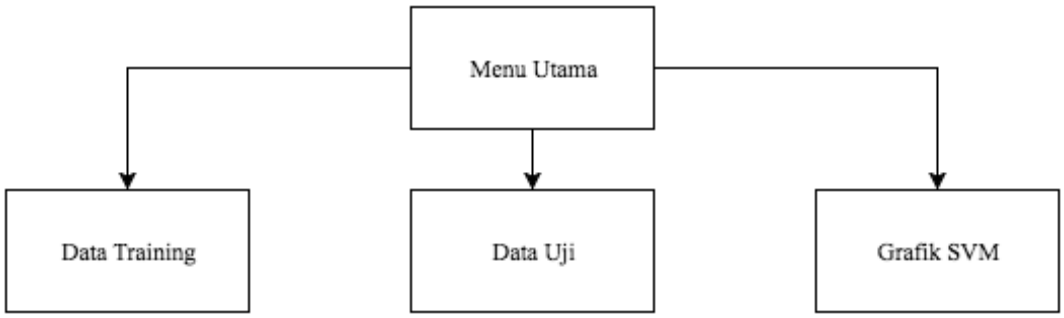
}



Gambar 3.19 Data Model Design

3.2.2 Perancangan Struktur Menu

Fungsi-fungsi yang dirancang pada tahap perancangan ini dibagi kedalam beberapa menu yang bertujuan untuk memudahkan pengoperasian program. Struktur menu ini dapat dilihat pada gambar 3.16.



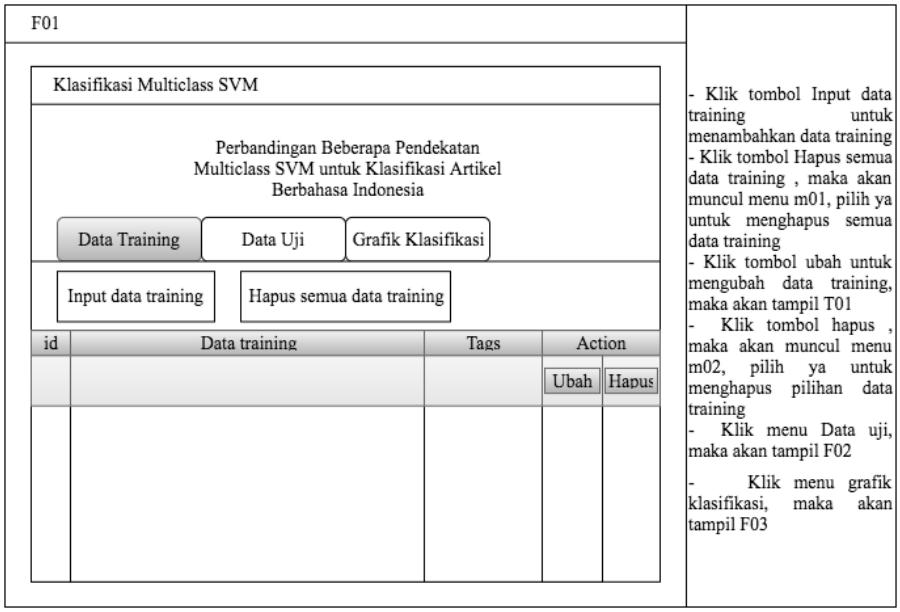
Gambar 3.20 Perancangan Menu

3.2.3 Perancangan Antarmuka

Perancangan antarmuka merupakan rancangan untuk menentukan tampilan dari sistem yang akan dibuat. Berikut adalah perancangan pada aplikasi yang akan dibangun.

1. F01 – *Form* Halaman Utama dan Pengolahan data *training*

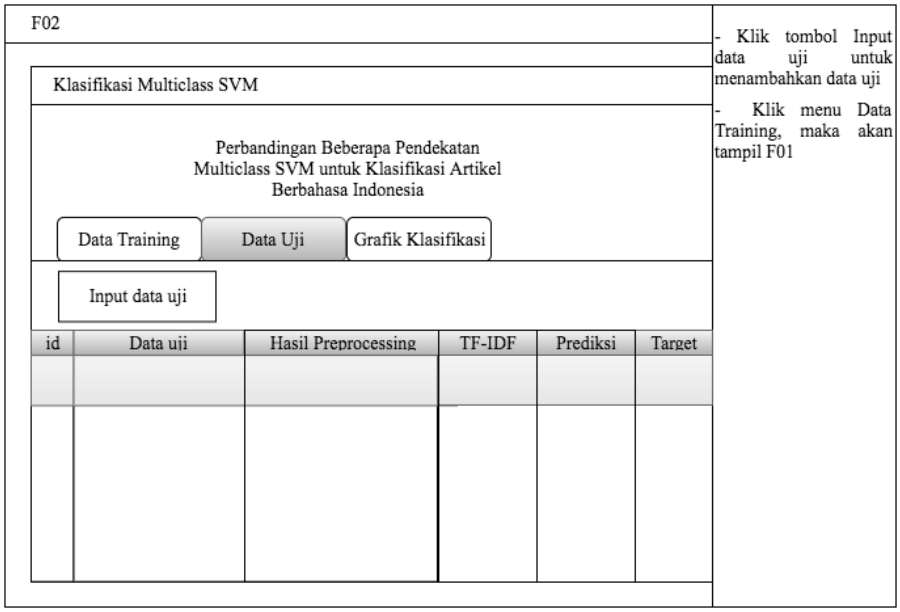
   Pada perancangan F01 adalah perancangan halaman utama pada saat pertama kali menjalankan aplikasi. Pada perancangan halaman ini juga ditampilkan input data *training*, hapus semua data training, hapus salah satu data training, dan ubah data training, namun menu data uji baru bisa digunakan setelah menginputkan data *training*.



Gambar 3.21 Perancangan Halaman Utama dan Data *Training*

2. F02 – *Form* Mengklasifikasi Data Uji

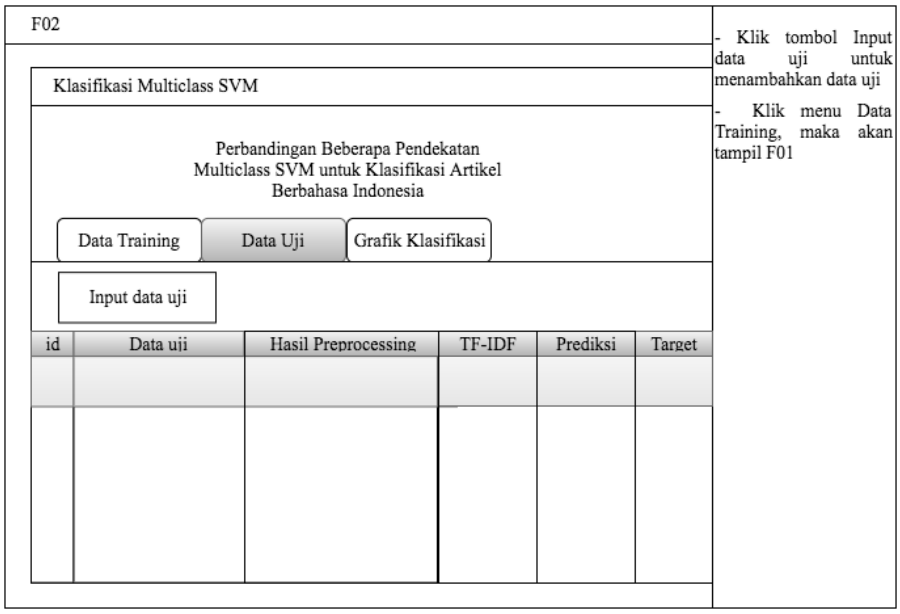
Pada perancangan F02 adalah perancangan untuk menampilkan data uji, hasil preprocessing,hasil pembobotan TF-IDF yang telah dilakukan pada data uji, serta kelas prediksi dan kelas target.



Gambar 3.22 Perancangan antarmuka pengolahan data uji

3. F02 – *Form* Menampilkan hasil *preprocessing*

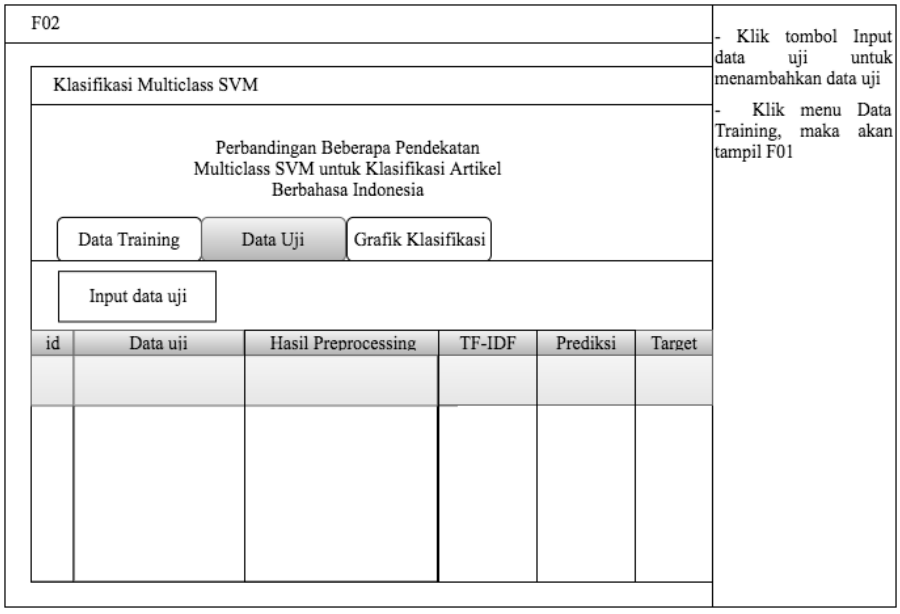
Menampilkan hasil preprocessing yang akan ditampilkan pada kolom hasil preprocessing.



Gambar 3.23 Perancangan antarmuka grafik klasifikasi

4. F02 – *Form* Menampilkan hasil prediksi

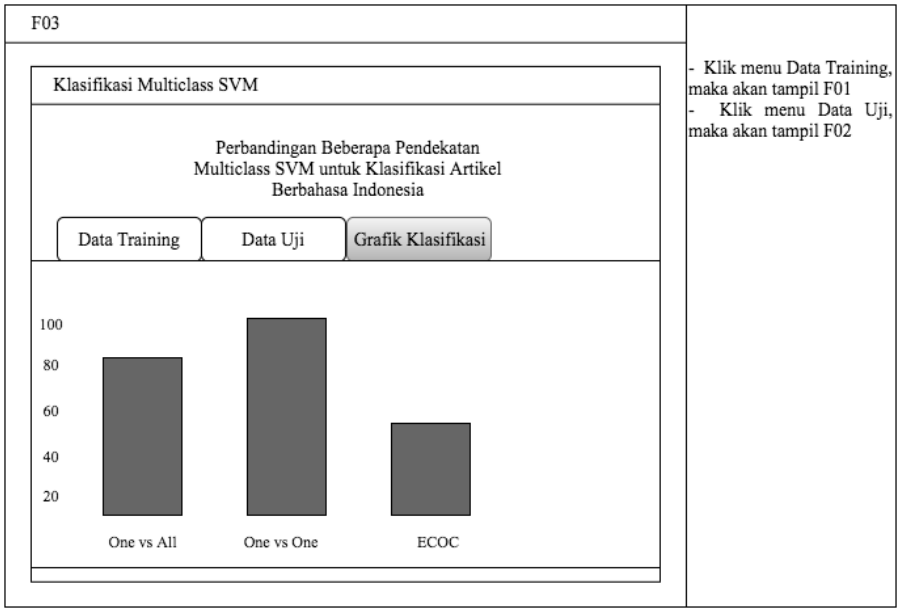
   Menampilkan hasil klasifikasi yang berupa prediksi kelas yang akan ditampilkan pada kolom prediksi berdasarkan metode *One vs All, One vs One* dan ECOC.



Gambar 3.24 Perancangan antarmuka prediksi

5. F03 – *Form* Menampilkan grafik klasifikasi

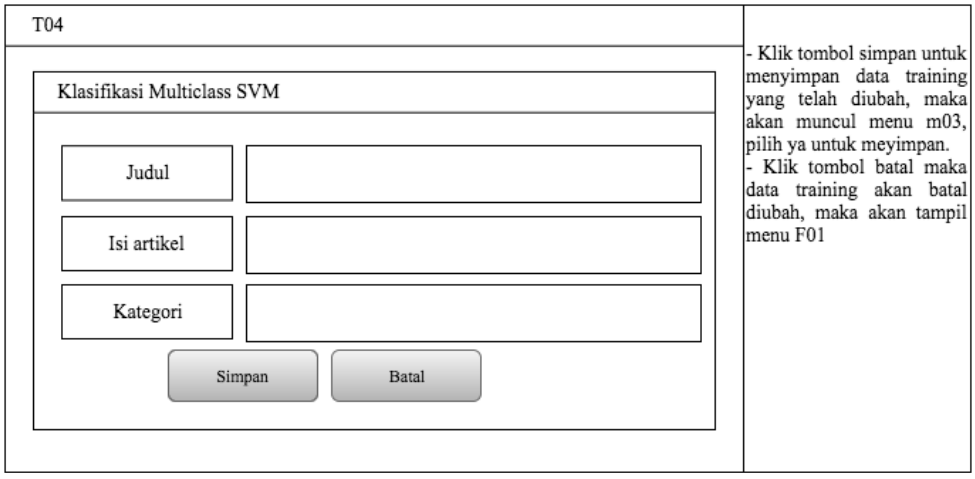
Pada perancangan F03 adalah perancangan untuk menampilkan grafik klasifikasi yang merupakan nilai *F1-Score* dari hasil klasifikasi.



Gambar 3.25 Perancangan antarmuka grafik klasifikasi

6. T01 - Perancangan antarmuka halaman ubah data *training*

Menampilkan halaman untuk melakukan perubahan data pada data training.

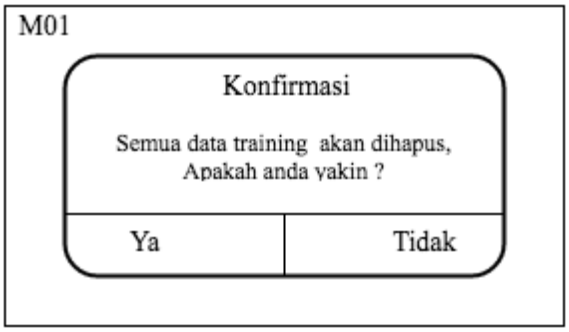


Gambar 3.26 Perancangan antarmuka ubah data

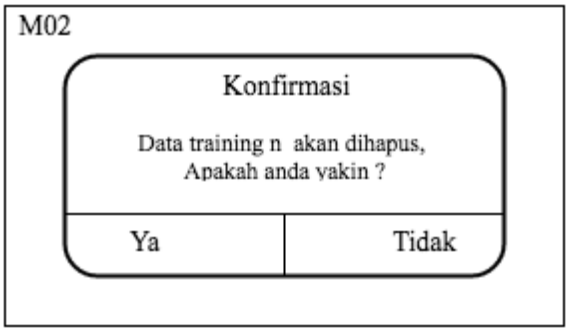
3.2.4 Perancangan Pesan

Tabel 3.41 Perancangan Pesan

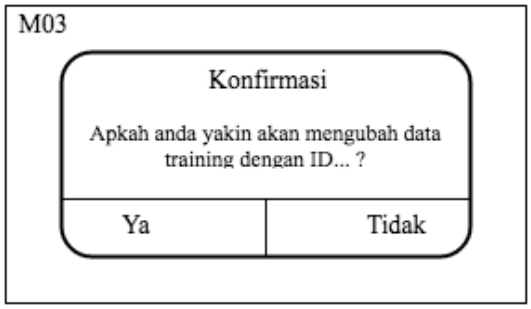
Kode Pesan	Keterangan Pesan
M01	Hapus semua data training
M02	Hapus data training
M03	Ubah data training
M04	Pesan gagal



Gambar 3.27 Perancangan Pesan M01

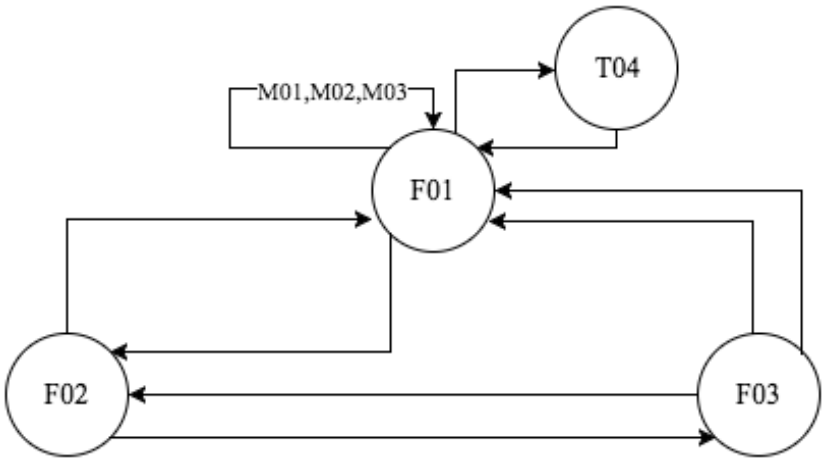


Gambar 3.28 Perancangan Pesan M02



Gambar 3.29 Perancangan Pesan M03

3.2.5 Perancangan Semantik



Gambar 3.30 Jaringan Semantik