



UNIVERSITAS
Dinamika

**RANCANG BANGUN SISTEM KLASIFIKASI KATEGORI BUKU
DENGAN ALGORITMA *NAÏVE BAYES CLASSIFIER* (STUDI KASUS :
PERPUSTAKAAN UNIVERSITAS DINAMIKA)**



TUGAS AKHIR

**Program Studi
S1 SISTEM INFORMASI**

UNIVERSITAS
Dinamika

Oleh :

CASMINA

17410100191

FAKULTAS TEKNOLOGI DAN INFORMATIKA

UNIVERSITAS DINAMIKA

2021

**RANCANG BANGUN SISTEM KLASIFIKASI KATEGORI BUKU
DENGAN ALGORITMA *NAÏVE BAYES CLASSIFIER* (STUDI KASUS :
PERPUSTAKAAN UNIVERSITAS DINAMIKA)**

TUGAS AKHIR

**Diajukan sebagai syarat untuk menyelesaikan
Program Sarjana**



UNIVERSITAS
Dinamika

Oleh:

Nama : CASMINA

NIM : 17410100191

Program Studi : S1 Sistem Informasi

**FAKULTAS TEKNOLOGI DAN INFORMATIKA
UNIVERSITAS DINAMIKA**

2021

TUGAS AKHIR

RANCANG BANGUN SISTEM KLASIFIKASI KATEGORI BUKU DENGAN ALGORITMA *NAÏVE BAYES CLASSIFIER* (STUDI KASUS : PERPUSTAKAAN UNIVERSITAS DINAMIKA)

Dipersiapkan dan disusun oleh :

Casmina

NIM : 17410100191

Telah diperiksa, diuji, dan disetujui oleh Dewan Penguji

Pada : 9 Februari 2021

Susunan Dewan Penguji

Pembimbing

I. Vivine Nurcahyawati, M.Kom.

NIDN. 0723018101

II. Julianto Lemantara, S.Kom., M.Eng.

NIDN. 0722108601



Digitally signed by Universitas
Dinamika
Date: 2021.02.09
16:18:48 +07'00'



Digitally signed by Julianto
Lemantara
DN: cn=Julianto Lemantara,
o=Universitas Dinamika, ou=S1
Sistem Informasi,
email=julianto@dinamika.ac.id, c=ID
Date: 2021.02.09 22:09:44 +07'00'

Pembahas

Tri Sagirani, S.Kom., M.MT.

NIDN. 0731017601



Digitally signed by
Universitas Dinamika
Date: 2021.02.10
09:21:29 +07'00'

Tugas Akhir ini telah diterima sebagai salah satu persyaratan

Untuk memperoleh gelar Sarjana



Dr. Jusak

Digitally signed by
Universitas Dinamika
Date: 2021.02.11
08:48:16 +07'00'

NIDN. 0708017101

Dekan Fakultas Teknologi dan Informatika

UNIVERSITAS DINAMIKA

SURAT PERNYATAAN
PERSETUJUAN PUBLIKASI DAN KERJASAMA KARYA ILMIAH

Sebagai mahasiswa Universitas Dinamika, saya:

Nama : Casmira

NIM : 17410100191

Program Studi : S1 Sistem Informasi

Fakultas : Fakultas Teknologi dan Informatika

Jenis Karya : Laporan Tugas Akhir

Judul Karya : **RANCANG BANGUN SISTEM KLASIFIKASI KATEGORI BUKU DENGAN ALGORITMA *NAÏVE BAYES CLASSIFIER* (STUDI KASUS : PERPUSTAKAAN UNIVERSITAS DINAMIKA)**

Menyatakan dengan sesungguhnya bahwa:

1. Demi pengembangan Ilmu Pengetahuan, Teknologi dan Seni, saya menyetujui memberikan kepada Universitas Dinamika Hak Bebas Royalti Non-Eksklusif (Non-Exclusive Royalti Free Right) atas seluruh isi/ sebagian karya ilmiah saya tersebut di atas untuk disimpan, diahlimediasikan dan dikelola dalam bentuk pangkalan data (*database*) untuk selanjutnya didistribusikan atau dipublikasikan demi kepentingan akademis dengan tetap mencantumkan nama saya sebagai penulis atau pencipta dan sebagai pemilik Hak Cipta.
2. Karya tersebut di atas adalah karya asli saya, bukan plagiat baik sebagian maupun keseluruhan, Kutipan karya atau pendapat orang lain yang ada dalam karya ilmiah ini adalah semata hanya rujukan yang dicantumkan dalam Daftar Pustaka saya.
3. Apabila dikemudian hari ditemukan dan terbukti terdapat tindakan plagiat pada karya ilmiah ini, maka saya bersedia untuk menerima pencabut terhadap gelar kerjasama yang telah diberikan kepada saya.

Demikian surat pernyataan ini saya buat dengan sebenarnya.

Surabaya, 9 Februari 2021

Yang Menyatakan



Casmira

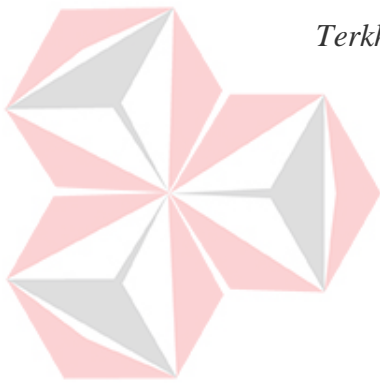


“Pray Love and Easy”

UNIVERSITAS
Dinamika

Saya persembahkan kepada kedua Orang Tua, Saudara, Teman yang memberi dukungan.

Terkhusus diri sendiri, terimakasih sudah berusaha dan lillah.



UNIVERSITAS
Dinamika

ABSTRAK

Perpustakaan Universitas Dinamika (Undika) mempunyai tempat yang nyaman dan fasilitas lengkap untuk kebutuhan pemustaka. Pada segi kemampuan teknologi Perpustakaan Undika mempunyai *Online Catalog* untuk melakukan pencarian buku. untuk penentuan kategori/kelas oleh petugas pengelola pustaka yang mengacu pada buku DDC23 mulai dari pencarian pada kelas utama ke kelas kedua dan kelas ketiga sebagai kelas spesifik atau kelas kecil dari penentuan kategori tersebut yang membutuhkan waktu kurang lebih 3 sampai 10 menit. Permasalahan yang ada yaitu pengelola pustaka kesulitan menentukan kategori yang sesuai karena penentuan kategori kurang cepat, dilakukan berdasarkan klasifikasi (nomor) yang ada pada DDC dari mulai mencari pada klasifikasi besar hingga pencarian spesifik ke klasifikasi kecil sampai menemukan kategori yang sesuai. Berdasarkan permasalahan tersebut, maka tujuan penelitian ini adalah untuk menghasilkan sistem klasifikasi kategori buku dengan algoritma *Naïve Bayes Classifier* (NBC) yang dapat melakukan klasifikasi kategori buku lebih cepat dan hasilnya tepat. Untuk merancang bangun sistem klasifikasi kategori buku dengan algoritma NBC menggunakan kerangka kerja Waterfall. NBC merupakan metode yang digunakan untuk klasifikasi yang mana prosesnya dibagi menjadi dua tahap yaitu tahap pembelajaran dan tahap pengujian. Klasifikasi menggunakan kelas 000 dari DDC23 yang mana datanya terbagi menjadi 311 data *Training* dan 78 data *Testing* dan hasil evaluasi pengujian akurasi yaitu sebesar 79% yang menunjukkan bahwa algoritma NBC dapat digunakan sebagai algoritma untuk pengklasifikasian kategori buku karena tingkat akurasinya yang cukup tinggi dan tingkat akurasi pengujian dipengaruhi oleh banyaknya jumlah data *Training* dan kamus kategori yang digunakan. Klasifikasi kategori dilakukan dalam waktu yang lebih cepat dengan sistem hanya membutuhkan 2 detik 70 milidetik.

Kata Kunci: *Klasifikasi teks, Naïve Bayes Classifier, Text preprocessing, Kategori buku.*

KATA PENGANTAR

Puji dan syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa, karena hanya atas berkat dan rahmat-Nya, Laporan Tugas Akhir yang berjudul Rancang Bangun Sistem Klasifikasi Kategori Buku dengan Algoritma NBC (Studi Kasus : Perpustakaan Universitas Dinamika) dapat diselesaikan dengan baik dan tepat waktu.

Dalam penyusunan Laporan Tugas Akhir ini penulis menyadari adanya banyak kekurangan dan keterbatasan maka melalui kesempatan yang sangat berharga ini Penulis menyampaikan ucapan terimakasih kepada semua pihak yang telah membantu dalam penyelesaian Laporan Tugas Akhir ini kepada:

1. Kedua orang tua penulis yang selalu mendoakan dan memberi kasih sayang yang tak terhingga.
2. Bapak Prof. Dr. Budi Jatmiko, M.Pd selaku Rektor Universitas Dinamika Surabaya.
3. Dr. Anjik Sukmaaji, S.Kom., M.Eng. selaku Ketua Program Studi S1 Sistem Informasi, Universitas Dinamika Surabaya.
4. Ibu Vivine Nurcahyawati, M.Kom. dan Bapak Julianto Lemantara, S.Kom., M.Eng. selaku Dosen pembimbing yang telah membimbing dengan sabar dan memberikan pengarahan dalam menyelesaikan Laporan Tugas Akhir ini.
5. Ibu Tri Sagirani, S.Kom., M.MT. selaku dosen penguji dalam Tugas Akhir ini.
6. Ibu Deasy Kumalawati, S.Pd. dan Bapak Totok Kariono, S.Sos. selaku Kepala dan bagian Perpustakaan Universitas Dinamika.
7. Pihak – pihak lain yang tidak dapat disebutkan satu persatu yang telah memberikan dukungan kepada penulis.

Penulis menyadari bahwa laporan ini masih jauh dari kesempurnaan. Oleh karena itu, penulis memohon kritik dan saran dari semua pihak sangat diharapkan untuk perbaikan di masa mendatang. Semoga Tuhan Yang Maha Esa memberikan imbalan yang setimpal atas segala bantuan yang diberikan.

Surabaya, 9 Februari 2021

Penulis

DAFTAR ISI

	Halaman
ABSTRAK	viii
KATA PENGANTAR	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xvi
BAB 1. PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	3
1.4 Tujuan	4
1.5 Manfaat	4
BAB 2. LANDASAN TEORI	5
2.1 Penelitian Terdahulu.....	5
2.2 Data Mining.....	7
2.3 Klasifikasi	7
2.4 Algoritma NBC	9
2.5 <i>Text Teks preprocessing</i>	11
2.6 Github Sastrawi	14
BAB 3. METODOLOGI PENELITIAN	15
3.1 Model SDLC	15
3.1.1 Communication	15
3.1.2 <i>Planning</i>	16
3.1.3 <i>Modeling</i>	16

3.1.4	<i>Construction</i>	36
3.1.5	<i>Deployment</i>	38
BAB 4. HASIL DAN PEMBAHASAN		42
4.1	Implementasi Sistem.....	42
4.1.1	Mengelola data <i>Training</i>	42
4.1.2	Testing.....	43
4.1.3	Klasifikasi Kategori.....	43
4.2	Pengujian Sistem	45
4.2.1.	Proses klasifikasi data <i>Training</i>	45
4.2.2.	<i>Testing</i>	45
4.3	Hasil Pengujian.....	46
4.4	Evaluasi	49
BAB 5. PENUTUP		50
5.1	Kesimpulan.....	50
5.2	Saran	50
DAFTAR PUSTAKA		51
DAFTAR RIWAYAT HIDUP		52
LAMPIRAN		53

DAFTAR TABEL

	Halaman
Tabel 2.1 Penelitian Terdahulu I	5
Tabel 2.2 Penelitian Terdahulu II	6
Tabel 3. 1 Struktur tabel kategori	33
Tabel 3. 2 Struktur tabel <i>Training</i>	33
Tabel 3. 3 Struktur tabel data testing	33
Tabel 3. 4 <i>Test Case</i>	38
Tabel 3. 5 hasil pengujian proses menyiapkan data <i>Training</i>	39
Tabel 3. 6 hasil pengujian <i>Testing</i>	39
Tabel 3. 7 hasil klasifikasi kategori Perpustakaan Undika	40
Tabel 4.1 hasil pengujian proses klasifikasi data <i>Training</i>	45
Tabel 4.2 hasil pengujian <i>Testing</i>	45
Tabel 4. 3 Hasil pengujian.....	46
Tabel L1.1 Perbandingan Algoritma I	53
Tabel L1.2 Perbandingan Algoritma II	54
Tabel L2. 1 Struktur tabel term	61
Tabel L2. 2 Struktur tabel data term2	61
Tabel L2. 3 Struktur tabel data training	61
Tabel L2. 4 Struktur tabel data term kategori	62
Tabel L2. 5 Struktur tabel data term hasil.....	62
Tabel L2. 6 Struktur tabel data NBC	63
Tabel L2. 7 Struktur tabel data probabilitas	63
Tabel L2. 8 Struktur tabel data klasifikasi	64
Tabel L2. 9 Struktur tabel data stopwords	64
Tabel L2. 10 Struktur tabel data kamus	64
Tabel L2.11 Hasil pengujian melihat data kategori.....	69
Tabel L2.12 hasil pengujian mengelola <i>filtering</i> tambah data	69
Tabel L2.13 hasil pengujian mengelola <i>filtering</i> hapus data	69
Tabel L2.14 hasil pengujian mengelola data <i>Training</i> tambah data.....	69
Tabel L2.15 hasil pengujian mengelola data <i>Training</i> hapus data	69
Tabel L2.16 hasil pengujian mengelola data <i>Training</i> ubah data.....	70

Tabel L2.17 hasil pengujian mengelola data klasifikasi	70
Tabel L2.18 hasil klasifikasi kategori Perpustakaan Undika	71
Tabel L3.1 Hasil pengujian melihat data kategori.....	94
Tabel L3.2 hasil pengujian mengelola Filtering tambah data	94
Tabel L3.3 hasil pengujian mengelola <i>Filtering</i> hapus data	95
Tabel L3.4 hasil pengujian mengelola data <i>Training</i> tambah data.....	95
Tabel L3.5 hasil pengujian mengelola data <i>Training</i> hapus data.....	95
Tabel L3.6 hasil pengujian mengelola data <i>Training</i> ubah data	95
Tabel L3.7 hasil pengujian mengelola data klasifikasi	95

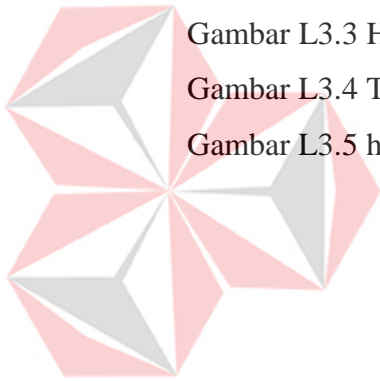


UNIVERSITAS
Dinamika

DAFTAR GAMBAR

	Halaman
Gambar 2.1 <i>Tokenizing Process</i>	12
Gambar 2. 2 <i>Filtering Process</i>	13
Gambar 2. 3 <i>Stemming Process</i>	14
Gambar 3.1 SDLC model <i>Waterfall</i> Menurut (Pressman & Maxim, 2015)	15
Gambar 3.2 IPO Diagram.....	18
Gambar 3.3 Skema Pengembangan	21
Gambar 3.4 Alur Proses Klasifikasi Naive Bayes	23
Gambar 3.5 Diagram tahap training	24
Gambar 3. 6 Diagram tahap testing klasifikasi	25
Gambar 3. 7 <i>System Flow</i> Mengelola data <i>Training</i>	26
Gambar 3. 8 <i>System Flow Testing</i>	27
Gambar 3. 9 <i>Context Diagram</i> sistem.....	28
Gambar 3. 10 <i>Data Flow Diagram</i> level 0.....	29
Gambar 3. 11 <i>Data Flow Diagram</i> level 1 proses menyiapkan data training.....	30
Gambar 3. 12 <i>Data Flow Diagram</i> level 1 klasifikasi kategori	30
Gambar 3. 13 <i>Conceptual data model</i> Sistem.....	31
Gambar 3. 14 <i>Physical data model</i> Sistem	32
Gambar 3. 15 <i>Interface</i> data <i>training</i>	34
Gambar 3. 16 <i>Interface Testing</i>	35
Gambar 3. 17 <i>Interface</i> klasifikasi kategori	35
Gambar 3. 18 <i>Interface</i> detil klasifikasi kategori	36
Gambar 3. 19 proses <i>Case folding</i>	36
Gambar 3. 20 proses <i>Tokenizing</i>	37
Gambar 3. 21 proses <i>filtering</i>	37
Gambar 3. 22 proses <i>Stemming</i>	37
Gambar 3. 23 halaman detil klasifikasi kategori	44
Gambar 4.1 Halaman Data <i>Training</i>	42
Gambar 4.2 halaman <i>Testing</i>	43
Gambar 4. 3 halaman klasifikasi Kategori	43

Gambar L2.1 <i>System Flow</i> Kategori.....	55
Gambar L2.2 <i>System Flow</i> mengelola data filtering.....	56
Gambar L2. 3 <i>System Flow</i> menyiapkan data training	57
Gambar L2. 4 <i>System Flow</i> klasifikasi kategori.....	58
Gambar L2.5 <i>System Flow</i> Mengelola data Klasifikasi.....	59
Gambar L2. 6 <i>System Flow</i> NBC	60
Gambar L2.7 <i>Interface</i> Beranda	65
Gambar L2.8 <i>Interface</i> About.....	66
Gambar L2.9 <i>Interface</i> Kategori	66
Gambar L2.10 <i>Interface</i> Filtering input data	67
Gambar L2.11 <i>Interface</i> Data klasifikasi	68
Gambar L3.1 Halaman Beranda	90
Gambar L3.2 Halaman Tentang	91
Gambar L3.3 Halaman Daftar Data Kategori	92
Gambar L3.4 Tampilan Input data <i>Filtering</i>	93
Gambar L3.5 halaman mengelola data klasifikasi.....	94



UNIVERSITAS
Dinamika

DAFTAR LAMPIRAN

	Halaman
Lampiran 1 Landasan Teori	53
Lampiran 2 Metodologi Penelitian	55
Lampiran 3 Pembahasan dan Hasil.....	90
Lampiran 4 Perhitungan manual NBC.....	96
Lampiran 5 hasil <i>Google form</i> kepuasan penerimaan sistem.....	102



UNIVERSITAS
Dinamika

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perpustakaan adalah tempat dimana banyaknya koleksi pustaka disimpan dan disediakan untuk pemustaka dengan pelayanan dan pengelolaan pustaka yang baik. Seperti dalam penelitian ini yaitu Perpustakaan Universitas Dinamika (Undika) mempunyai tempat yang nyaman dan fasilitas lengkap untuk kebutuhan pemustaka. Dalam menjalankan operasi pelayanan tentu saja sudah tersusun struktur organisasi sebagai pembagian tugas masing-masing. Banyak koleksi buku dan tempat baca yang menyenangkan untuk mendukung suasana baca pemustaka, dimana pada saat ini Perpustakaan Undika memiliki koleksi buku yaitu 35.966 eksemplar dan saat ini total buku yang aktif sebanyak 26.267 eksemplar, data yang ada dalam katalog *Undika Library* sebanyak 14.658 koleksi buku, sementara itu total koleksi (Buku, Majalah, CD/AV, Karya Ilmiah) Perpustakaan Undika yaitu sejumlah 49.525. Penurunan jumlah koleksi buku ini dikarenakan kejadian seperti hilang, digudangkan, dan rusak. Perpustakaan Undika sudah menjadi anggota dari *Microsoft Campus Agreement* dan mempunyai fasilitas baru yaitu *e-Resources Center* berisi koleksi dalam bentuk digital yang dapat di *download* khusus untuk sivitas akademika Undika. Fasilitas lain yang terdapat pada Perpustakaan Undika yaitu Komputer Katalog, Informasi Koleksi Baru, Layanan Loker, *e-Resources Center*, *Hotspot Wireless Fidelity* (Wifi), *Integrated Learning Terminal* (ILT).

Pada segi kemampuan teknologi Perpustakaan Undika tentu saja memberikan layanan yang dapat memudahkan sivitas Undika dan pengguna umum dalam melakukan pencarian koleksi buku dimanapun dan kapanpun yaitu dengan *Online Catalog*. User yang menggunakan sistem *Online Catalog* yaitu pengelola pustaka dan Pemustaka yaitu pengguna perpustakaan, yaitu perseorangan, kelompok orang, masyarakat, atau lembaga yang memanfaatkan fasilitas layanan Perpustakaan Undika. Data buku yang ada pada sistem *Online Catalog* sejumlah 14.658 koleksi/judul. Prosesnya berawal dari pengolahan untuk mendaftarkan koleksi yang akan diolah ke dalam aplikasi pengadaan. Proses mendaftarkan koleksi ke aplikasi pengadaan ini dilakukan oleh kepala bagian

perpustakaan. Kemudian penentuan kategori/kelas oleh petugas pengelola pustaka yang mengacu pada buku DDC23 mulai dari pencarian pada kelas utama ke kelas kedua dan kelas ketiga sebagai kelas spesifik atau kelas kecil dari penentuan kategori tersebut. Setelah menemukan kategori yang sesuai kemudian pengelola melakukan penambahan buku baru untuk *Online Catalog* menggunakan aplikasi desktop yaitu Aplikasi Pengadaan untuk pembelian buku setelah selesai akan dilanjutkan ke pengolahan untuk menentukan data dengan pengindukan atau penentuan klasifikasi yang membutuhkan waktu kurang lebih 3 sampai 10 menit dilanjutkan input deskripsi, setelah klik simpan maka data buku akan masuk ke sistem *Online Catalog* dan dapat diakses secara umum dengan cara mengakses *Undika Library* lalu memilih *Online Catalog*. Sementara itu untuk melakukan pencarian koleksi, pemustaka dapat memilih jenis koleksi seperti Buku, Majalah, *Software*, atau Tugas Akhir kemudian memilih kriteria pencariannya seperti Judul, Pengarang, Subyek, atau ISBN dan memasukan kata kunci pencarian.

Permasalahan yang muncul yaitu pengelola pustaka kesulitan menentukan kategori yang sesuai karena penentuan kategori kurang cepat, dilakukan berdasarkan klasifikasi (nomor) yang ada pada DDC dari mulai mencari pada klasifikasi besar hingga pencarian spesifik ke klasifikasi kecil sampai menemukan kategori yang sesuai. Dampak yang ditimbulkan yaitu penentuan kategori memakan waktu dan memungkinkan terjadinya kesalahan pengklasifikasian kategori buku yang akan mempengaruhi pencarian katalog pada saat digunakan oleh pemustaka.

Solusi yang digunakan untuk membantu mempermudah pengelola dalam melakukan klasifikasi kategori yang sesuai yaitu dengan merancang sistem klasifikasi kategori buku dengan algoritma NBC (Studi Kasus : Perpustakaan Undika) untuk penentuan klasifikasi secara cepat dan meningkatkan kesesuaian kategori buku. “NBC merupakan salah satu machine learning yang menggunakan perhitungan probabilitas. Proses klasifikasi dibagi menjadi dua fase yaitu *learning/training* dan *testing/classify*” (Ginting & Trinanda, 2013). NBC mempunyai nilai akurasi tinggi dibuktikan oleh 2 penelitian yang melakukan perbandingan algoritma untuk model klasifikasi yang ditunjukan pada lampiran 1 tabel 2.3 dan 2.4 dibawah. Kriteria yang digunakan ada 16 meliputi *Computer Program Language*, *Computer Programs*, *Operating Systems*, *Computers*,

Computer Network, Computer Security, Computer Science, Internet, Computer Graphics, Library, Operations research, Microprocessors, Artificial Intelligence, System Analysis, Data Communication, Microcomputers. Dari kriteria yang dipilih berasal dari kelas 000 DDC23 merupakan kelas yang memiliki jumlah total terbanyak yaitu sejumlah 11691 dari total judul buku yang ada yaitu sejumlah 14.658. kemudian kriteria tersebut adalah perwakilan kriteria yang memiliki jumlah data paling banyak dari total kategori yang ada pada kelas 000.

1.2 Rumusan Masalah

Bagaimana merancang Bangun Sistem Klasifikasi Kategori Buku dengan Algoritma NBC (Studi Kasus : Perpustakaan Undika) untuk penentuan klasifikasi secara cepat dan meningkatkan kesesuaian kategori buku.

1.3 Batasan Masalah

Berdasarkan rumusan masalah di atas, batasan masalah dalam tugas akhir sebagai berikut:

1. Data klasifikasi yang digunakan hanya dari buku yang menggunakan bahasa Indonesia.
2. Kriteria kategori menggunakan klasifikasi kelas kecil dari kelas utama/besar DDC23.
3. Klasifikasi kategori menggunakan NBC.
4. Data yang digunakan untuk klasifikasi kategori buku berasal dari koleksi Perpustakaan Undika.
5. Kriteria kategorinya ada 16 macam yaitu *Computer Program Language, Computer Programs, Operating Systems, Computers, Computer Network, Computer Security, Computer Science, Internet, Computer Graphics, Library, Operations research, Microprocessors, Artificial Intelligence, System Analysis, Data Communication, Microcomputers.*

1.4 Tujuan

Berdasarkan uraian diatas tujuan penelitian yang dibuat yaitu :
Menghasilkan sistem klasifikasi kategori buku dengan algoritma NBC (Studi Kasus : Perpustakaan Undika) untuk penentuan klasifikasi secara cepat dan meningkatkan kesesuaian kategori buku.

1.5 Manfaat

Berikut manfaat dari pengembangan sistem yang akan dibuat :

1. Memudahkan pengelola pustaka pada saat pengklasifikasian kategori buku secara cepat dan meningkatkan kesesuaian hasil klasifikasi.
2. Membantu pengembangan sistem klasifikasi kategori untuk meringankan kinerja dan meningkatkan pelayanan perpustakaan.



UNIVERSITAS
Dinamika

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Berikut adalah data dari penelitian terdahulu yang menjadi perbandingan dengan penelitian yang akan dilakukan.

Tabel 2.1 Penelitian Terdahulu I

No.	Bagian	Isi
1.	Judul	Klasifikasi Jenis Buku Berdasarkan Judul Dan Sinopsis Menggunakan Metode NBC <i>Classifier</i> (Studi Kasus : STMIK Kadiri)
2.	Pengarang	Bonifacius Vicky Indriyono, Ema Utami, dan Andi Sunyoto
3.	Tahun	2015
4.	Data yang digunakan	Buku fisik diambil berasal dari 9 kategori dimana masing-masing kategori diambil 100 buah buku.
5.	Klasifikasi	Klasifikasinya ada 9 yaitu <i>Database, MS.Office, Pemrograman web, Delphi, DesainGrafis, java , Internet, SPSS, 3DsMax.</i>
6.	Akurasi	Pada skenario kedua menunjukkan akurasi sbb : ketika Jumlah Dokumen <i>Training</i> 90 dan Jumlah Dokumen <i>Testing</i> 10 maka akurasinya 90%. ketika Jumlah Dokumen <i>Training</i> 30 dan Jumlah Dokumen <i>Testing</i> 70 maka akurasinya 74,2%.
7.	Hasil	Akurasi mencapai nilai prosentase tinggi dengan semakin banyaknya penggunaan dokumen sebagai dokumen <i>training</i> dan sebaliknya akurasi akan menurun dengan semakin sedikitnya penggunaan dokumen <i>training</i> .
8.	Kesimpulan	<ol style="list-style-type: none"> 1. Metode NBC dan <i>Porter Stemmer</i> dapat diimplementasikan dalam sebuah <i>prototype</i> aplikasi klasifikasi jenisbuku. Aplikasi ini dibangun dan dilengkapi dengan aturan-aturan sesuai dengan tahap-tahap dalam NBC dan aturan-aturan <i>stemming</i> menurut <i>porterstemmer</i> untuk bahasa Indonesia (<i>Stemmer Tala</i>). 2. <i>Porter Stemmer</i> untuk bahasa Indonesia memiliki pengaruh yang signifikan terhadap proses klasifikasi. Hal ini dikarenakan algoritma <i>stemmer</i> ini dapat digunakan untuk membentuk kata dasar yang sebenarnya sehingga memudahkan dan mempercepat dalam proses perbandingan antara kata kunci dengan teks dokumen uji. 3. Pada hasil uji coba skenario kedua dapat disimpulkan bahwa tingginya tingkat akurasi diperoleh dengan semakin banyaknya penggunaan dokumen sebagai data <i>training</i> dan sebaliknya akurasi akan menurun dengan semakin sedikitnya penggunaan data <i>training</i>.

Tabel 2.2 Penelitian Terdahulu II

No.	Bagian	Isi
1.	Judul	Klasifikasi Judul Buku dengan Algoritma <i>Naïve Bayes</i> dan Pencarian Buku pada Perpustakaan Jurusan Teknik Elektro
2.	Pengarang	Vidya Rizqiyani, Anggraini Mulwinda, dan Riana Defi Mahadji Putri
3.	Tahun	2017
4.	Data yang digunakan	Data buku berasal dari Perpustakaan Teknik Elektro. Data pelatihan 174 dan data pengujian 75
5.	Klasifikasi	Klasifikasinya ada 6 kategori yaitu Jaringan Komputer, pemrograman komputer, Pendidikan, Matematika, Fisika, Teknik elektro.
6.	Akurasi	Tingkat akurasi 97,78% yang menunjukkan nilai <i>excellent classification</i> .
7.	Hasil	Hasil klasifikasi sistem pada tahap pengujian melakukan kesalahan sebanyak 5 dokumen uji dari total keseluruhan sebanyak 75 dokumen uji. Kesalahan yang diperoleh yaitu 3 kesalahan pada dokumen uji kategori Jaringan Komputer, dan 2 kesalahan pada dokumen uji kategori Pemrograman Komputer. Sedangkan untuk hasil performance measure menunjukkan bahwa pengujian evaluasi sistem dalam pengklasifikasian buku Perpustakaan Jurusan Teknik Elektro dinyatakan sangat baik, valid, dan akurat ditunjukkan dengan hasil pengukuran mendapatkan nilai rata-rata <i>recall</i> 88,20%, <i>precision</i> 94,56%, <i>f-measure</i> 90,46%, dan akurasi 97,78% . Berdasarkan standar tingkat akurasi, maka hasil akurasi tersebut menunjukkan nilai <i>excellent classification</i> .
8.	Kesimpulan	<ol style="list-style-type: none"> 1. Sistem klasifikasi dan pencarian buku ini mempunyai 2 hak akses yaitu pengelola dan <i>user</i>. Sistem ini menggunakan algoritma <i>Naïve Bayes</i> dalam pengklasifikasian judul buku. 2. Klasifikasi dan pencarian buku perpustakaan Jurusan Teknik Elektro dilakukan dengan menggunakan 249 dokumen. Hasil pengujian algoritma NBC diperoleh 70 dokumen relevan dan 5 dokumen tidak relevan dengan nilai masing-masing <i>recall</i> sebesar 88,20%, <i>precision</i> sebesar 94,56%, <i>f-measure</i> sebesar 90,46%, dan akurasi sebesar 97,78% sehingga dapat disimpulkan bahwa algoritma NBC dapat digunakan dalam otomatisasi klasifikasi judul buku Perpustakaan Jurusan Teknik Elektro. 3. Pengujian <i>black box</i> pada sistem menghasilkan uji fungsi menu maupun tombol dapat digunakan dengan baik dan tidak didapatkan <i>error</i> pada saat pengujian.

Penelitian saat ini akan merancang sistem klasifikasi kategori buku dengan algoritma NBC (Studi Kasus : Perpustakaan Undika) untuk penentuan klasifikasi secara cepat dan meningkatkan kesesuaian kategori buku. kriteria klasifikasi

berasal dari kelas 000 berdasarkan DDC23. Klasifikasinya ada 16 kategori yaitu *Computer Program Language, Computer Programs, Operating Systems, Computers, Computer Network, Computer Security, Computer Science, Internet, Computer Graphics, Library, Operations research, Microprocessors, Artificial Intelligence, System Analysis, Data Communication, Microcomputers*. Dan data yang digunakan berasal dari koleksi Perpustakaan Undika. Jumlah data sample yang digunakan dihitung dengan rumus Slovin sebanyak 389 buku yang terdiri dari 80% data pelatihan dan 20% data pengujian.

2.2 Data Mining

Data Mining adalah proses menemukan pola yang menarik dari sejumlah besar data. Sebagai proses penemuan pengetahuan, biasanya melibatkan pembersihan data, integrasi data, pemilihan data, transformasi data, penemuan pola, evaluasi pola, dan penyajian pengetahuan. *Data Mining* dapat dilakukan pada semua jenis data asalkan data tersebut berguna untuk aplikasi target, seperti *database, data warehouse, transactional data*, dan *advanced data types*. Tipe data lanjutan termasuk data yang berhubungan dengan waktu atau sekuensing, aliran data, data spasial dan spasial, data teks dan multimedia, grafik dan data jaringan, dan data *Web*. Fungsi *data mining* digunakan untuk menentukan jenis pola atau diketahui yang dapat ditemukan dalam tugas *data mining*. *Fungsionalitas* tersebut meliputi karakterisasi dan diskriminasi, penambangan pola, asosiasi, dan korelasi yang sering, klasifikasi dan regresi, analisis cluster, dan deteksi outlier. Karena jenis data baru, aplikasi baru, dan permintaan analisis baru terus bermunculan, tidak diragukan lagi kita akan melihat semakin banyak tugas *data mining* baru di masa depan (Han, Kamber, & Pei, 2012).

2.3 Klasifikasi

Menurut Sulistyio Basuki (1991) dalam (Darmono, 2001) klasifikasi berasal dari kata Latin “*classis*” atau proses pengelompokan, artinya mengumpulkan benda/entitas yang sama serta memisahkan benda/entitas yang tidak sama. Sulistyio-Basuki mendefinisikan klasifikasi yang diterapkan di pusat informasi dan perpustakaan adalah penyusunan sistematis terhadap buku atau bahan pustaka lain

atau katalog atau entri indeks berdasarkan subjek, dalam cara yang paling berguna bagi mereka yang membaca atau mencari informasi.

Pada penelitian ini klasifikasi dilakukan dengan menggunakan DDC23 yang mana klasifikasinya diambil dari salah satu kelas besar berikut,

The Ten Main Classes

000 *Computer science, information & general works*

100 *Philosophy & psychology*

200 *Religion*

300 *Social sciences*

400 *Language*

500 *Science*

600 *Technology*

700 *Arts & recreation*

800 *Literature*

900 *History & geography*

Kemudian dari 9 kelas utama tersebut diatas dipilih satu kelas yaitu pada nomor 000 *Computer science, information & general works* kelas ini memiliki total buku sebanyak 11.691 dari total keseluruhan buku sebanyak 14.658. Dari satu kelas yang dipilih pada nomor 000 terdapat kelas detilnya yaitu,

000 *Computer science, information, generalworks*

SUMMARY

001–006 [*Knowledge, the book, systems, computer science*]

010 *Bibliography*

020 *Library and information sciences*

030 *General encyclopedic works*

050 *General serial publications*

060 *General organizations and museology*

070 *Documentary media, educational media, news media; journalism, publishing*

080 *General collections*

090 *Manuscripts, rare books, other rare printed materials*

Kemudian dari kelas kelas tersebut akan dilakukan pencarian dan penentuan ke kelas spesifik atau detilnya yang mana satu judul buku akan masuk dan diklasifikasikan dengan kategori tertentu.

2.4 Algoritma NBC

NBC merupakan salah satu metode *machine learning* yang memanfaatkan perhitungan probabilitas dan statistic yang dikemukakan oleh ilmuwan Inggris Thomas. *Bayes*, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Dasar dari *Naïve Bayes* yang dipakai dalam pemrograman adalah rumus *Bayes*:

$$P(A|B) = (P(B|A) * P(A))/P(B) \quad (1)$$

Peluang kejadian A sebagai B ditentukan dari peluang B saat A, peluang A, dan peluang B. Pada pengaplikasiannya nanti rumus ini berubah menjadi :

$$P(C_i|D) = (P(D|C_i) * P(C_i))/P(D) \quad (2)$$

NBC atau bisa disebut sebagai *Multinomial* NBC merupakan model penyederhanaan dari Metode Bayes yang cocok dalam pengklasifikasian teks atau dokumen. Persamaannya adalah:

$$VMAP = \arg \max P(V_j|a_1, a_2, \dots, a_n) \quad (3)$$

Menurut persamaan 3, maka persamaan 4 dapat ditulis:

$$VMAP = \arg \max_{v_j \in V} \frac{P(a_1, a_2, \dots, a_n|v_j)P(v_j)}{P(a_1, a_2, \dots, a_n)} \quad (4)$$

$P(a_1, a_2, \dots, a_n)$ konstan, sehingga dapat dihilangkan menjadi :

$$VMAP = \arg \max_{v_j \in V} P(a_1, a_2, \dots, a_n|v_j)P(v_j) \quad (5)$$

Karena $P(a_1, a_2, \dots, a_n | v_j)$ sulit untuk dihitung, maka akan diasumsikan bahwa setiap kata pada dokumen tidak mempunyai keterkaitan.

$$VMAP = \arg \max_{v_j \in V} (v_j) \prod_i P(a_i | v_j) \quad (6)$$

Keterangan :

$$(v_j) = \frac{|docs_j|}{|Contoh|} \quad (7)$$

$$(w_k | v_j) = \frac{n_{k+1}}{n + |kosakata|} \quad (8)$$

Dimana Untuk :

$P(v_j)$: Probabilitas setiap dokumen terhadap sekumpulan dokumen

$P(w_k | v_j)$: Probabilitas kemunculan kata w_k pada suatu dokumen dengan kategori $class_{v_j}$

$|docs|$: frekuensi dokumen pada setiap kategori

$|Contoh|$: jumlah dokumen yang ada.

N_k : frekuensi kata ke- k pada setiap kategori

Kosakata: jumlah kata pada dokumen test

Pada persamaan (8) terdapat suatu penambahan 1 pada pembilang, hal ini dilakukan untuk mengantisipasi jika terdapat suatu kata pada dokumen uji yang tidak ada pada setiap dokumen data *training*.

Cara Kerja NBC

Cara kerja NBC melalui dua tahapan, yaitu :

1. *Learning* (Pembelajaran)

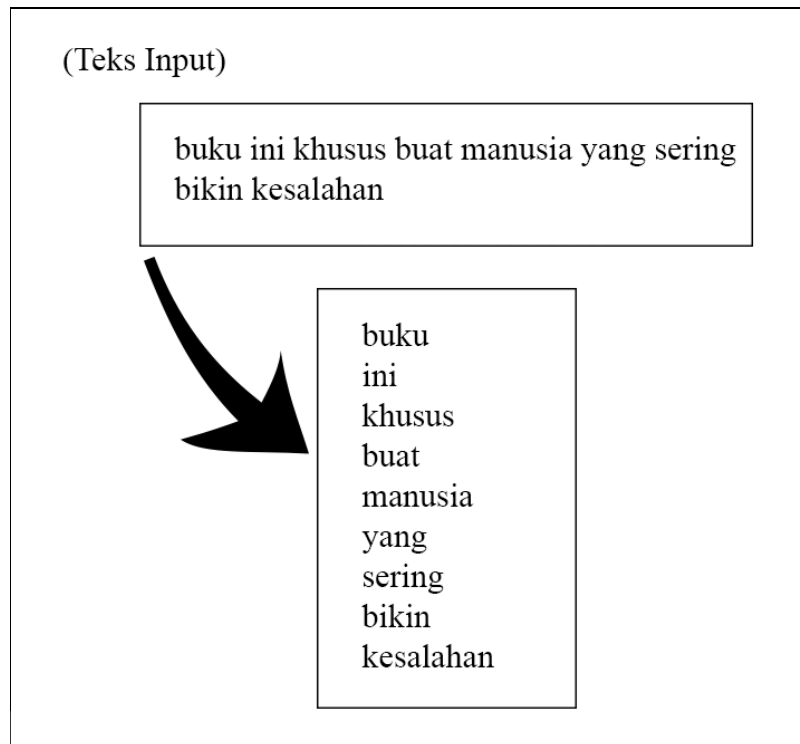
NBC adalah suatu metode yang termasuk ke dalam *supervised learning*, maka akan dibutuhkan pengetahuan awal untuk dapat mengambil keputusan. Langkah-langkah:

- a. Step 1 : Bentuk *vocabulary* pada setiap dokumen training
 - b. Step 2 : Hitung probabilitas pada setiap kategori $P(v_j)$.
 - c. Step 3 : Tentukan frekuensi setiap kata w_k pada setiap kategori $P(w_k|v_j)$
2. *Classify* (Pengklasifikasian). Langkah-langkahnya adalah :
- a. Step 1 : Hitung $P(v_j) \prod P(w_k | v_j)$ untuk setiap kategori.
 - b. Step 2 : Tentukan kategori dengan nilai $P(v_j) \prod P(w_k | v_j)$ maksimal (Kusuma & Srirahayu, 2016).

2.5 Text Teks preprocessing

Tahapan awal dari klasifikasi teks ialah *Text Teks preprocessing* yang bertujuan untuk mempersiapkan teks menjadi data yang akan mengalami pengolahan pada tahapan berikutnya. Beberapa contoh tindakan yang dapat dilakukan pada tahap ini, seperti proses *Tokenizing*, yaitu memecah suatu kalimat menjadi sekumpulan kata. Selain itu pada tahapan ini biasanya juga dilakukan *Case Folding*, yaitu pengubahan karakter huruf menjadi huruf kecil. Adapun penjelasan rinci mengenai *Case Folding* dan *Tokenizing* adalah:

1. *Case Folding* yaitu mengubah semua huruf dalam teks menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter.
2. *Tokenizing* adalah sebuah proses untuk memilah isi teks sehingga menjadi satuan kata-kata. Proses ini cukup rumit untuk sebuah program komputer karena beberapa karakter dapat dijadikan sebagai pembatas (delimiter) dari token-token itu sendiri. Contoh dari tahap ini dapat dilihat pada Gambar 1 dibawah ini (Setiawan, Astuti, & Kridalaksana, 2015).

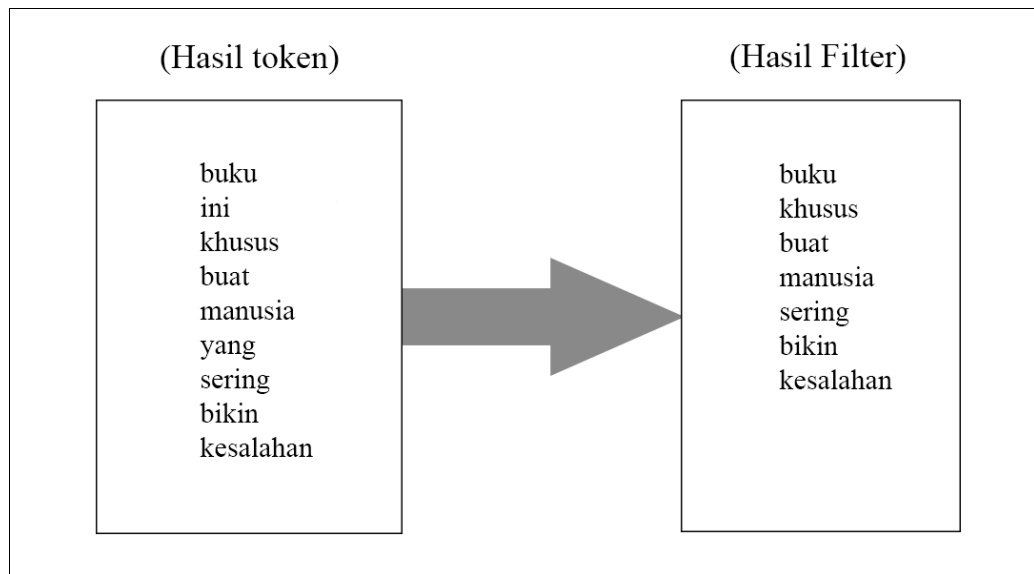


Gambar 2.1 *Tokenizing Process*

Dari proses Tokenizing seperti diperlihatkan pada gambar 2.1 diatas adalah proses penyusunan berdasarkan tiap kata pada kalimat.

3. *Filtering*

Tahapan *Filtering* merupakan tahap mengambil kata-kata penting dari hasil *token*. Bisa menggunakan algoritma *Stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). *Stoplist / Stopword* adalah kata-kata yang tidak *deskriptif* yang dapat dibuang dalam pendekatan *bag-of-words*. Contoh *Filtering* adalah “yang”, “dan”, “di”, “dari” dan seterusnya. Dengan menggunakan daftar *Stoplist*, maka setiap kata dalam koleksi akan dicocokkan dengan kata-kata yang ada.

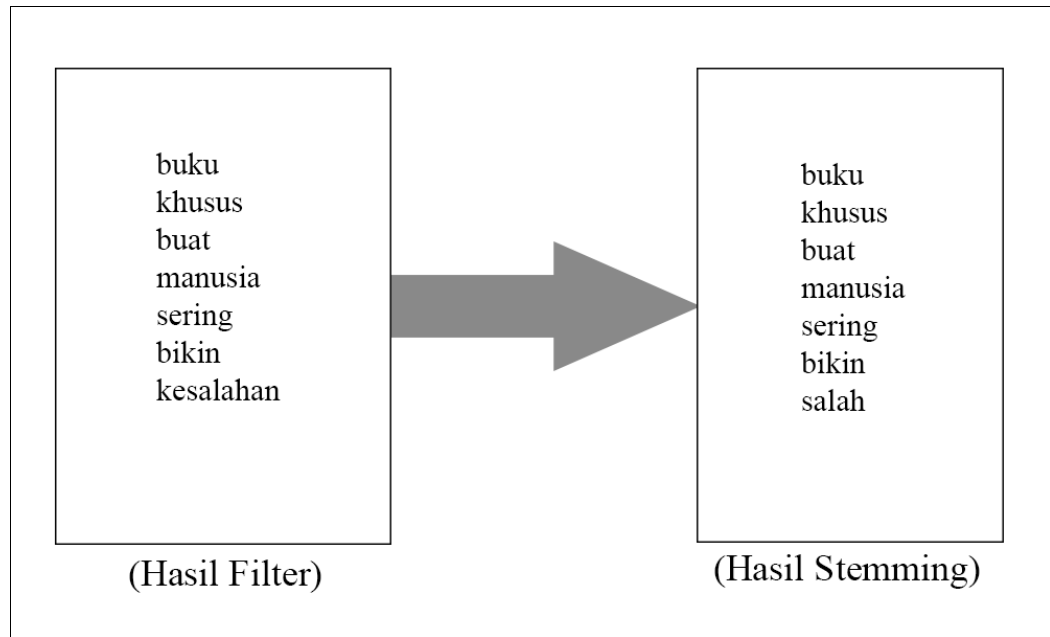


Gambar 2. 2 *Filtering Process*

Dari Gambar 2.2 diatas merupakan proses *Filtering* untuk mengambil kata-kata penting dari hasil *Tokenizing* dan membuang kata-kata yang tidak deskriptif.

4. *Stemming*

Stemming merupakan suatu proses untuk mereduksi kata ke bentuk dasarnya. *Stemming* dengan kata lain merupakan suatu proses yang menyediakan suatu pemetaan antara berbagai kata dengan morfologi yang berbeda menjadi satu bentuk dasar (stem). Tahap *stemming* merupakan tahap mencari akar (*root*) kata dari tiap kata hasil *filtering*. Contoh dari tahapan ini pada teks berbahasa Indonesia dapat dilihat pada Gambar 3 (Setiawan, Astuti, & Kridalaksana, 2015).



Gambar 2. 3 *Stemming Process*

Dari gambar 2.3 diatas merupakan Proses *Stemming* untuk pengelompokan kata dasar dari kata-kata yang memiliki imbuhan.

2.6 Github Sastrawi

GitHub adalah manajemen proyek dan sistem *versioning* code sekaligus *platform* jaringan sosial yang dirancang khusus bagi para developer. Lalu, apa saja fungsi GitHub? Dengan *platform* ini, Anda bisa bekerja bersama-sama dengan rekan dari berbagai belahan dunia, merencanakan proyek, dan bahkan *tracking* (melacak) pekerjaan Anda (Ariata, 2019).

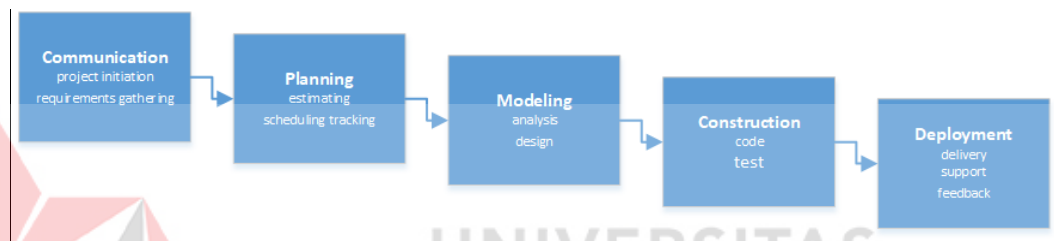
Sastrawi adalah pustaka PHP sederhana yang menyediakan *stemming* kata-kata dalam Bahasa Indonesia (Bahasa). Terlepas dari kesederhanaannya, perpustakaan ini dirancang dengan kualitas tinggi dan didokumentasikan dengan baik (Librian, 2017).

BAB III

METODOLOGI PENELITIAN

3.1 Model SDLC

Model air terjun, kadang-kadang disebut siklus hidup klasik, menyarankan pendekatan sistematis, sekuensial 2 untuk pengembangan perangkat lunak yang dimulai dengan spesifikasi kebutuhan pelanggan dan berkembang melalui perencanaan, pemodelan, konstruksi, dan penerapan, yang berpuncak pada dukungan berkelanjutan dari perangkat lunak yang sudah selesai (Pressman & Maxim, 2015).



Gambar 3.1 SDLC model *Waterfall* Menurut (Pressman & Maxim, 2015)

3.1.1 Communication

Pada tahap awal ini peneliti melakukan observasi pada Perpustakaan Undika dan melakukan wawancara langsung dengan pihak Perpustakaan Undika yaitu Ibu Deasy Kumalawati sebagai Kepala Bagian Perpustakaan dan Bapak Totok sebagai petugas pengelolaan pustaka. Wawancara dilakukan pada tanggal 7,8,19,21 Oktober 2020 di Perpustakaan Undika yang beralamat di Jl. Raya Kedung Baruk, 98 Kota Surabaya Jawa Timur, peneliti menganalisis data dan modul yang akan dikerjakan ditemukan permasalahan yang ada yaitu pengelola pustaka kesulitan menentukan kategori yang sesuai karena penentuan kategori kurang cepat. Klasifikasi dilakukan berdasarkan kelas yang ada pada DDC dari mulai mencari pada klasifikasi besar hingga pencarian ke klasifikasi kecil sampai menemukan kategori yang sesuai. Dampak yang ditimbulkan yaitu penentuan kategori memakan waktu dan memungkinkan terjadinya pengklasifikasian kategori buku tidak sesuai yang akan mempengaruhi pencarian katalog pada saat digunakan oleh pemustaka.

3.1.2 *Planning*

Pada tahap ini akan dilakukan penjadwalan untuk membuat rincian waktu pertahap yang akan dikerjakan. Untuk gambaran yang lebih jelas akan ditunjukkan pada Jadwal Kerja.

3.1.3 *Modeling*

Tahap ini digunakan untuk mengubah kebutuhan-kebutuhan diatas menjadi representasi sebelum *coding* dimulai. Mengimplementasikan kebutuhan yang telah disebutkan pada tahap sebelumnya. Seperti 2 aktivitas sebelumnya, maka proses ini juga akan didokumentasikan sebagai konfigurasi dari Aplikasi, pengerjaan yang akan dilakukan meliputi : Analisis proses bisnis, Analisis kebutuhan fungsional dan nonfungsional. Setelah melakukan analisis maka akan dilanjutkan dengan pembuatan diagram IPO dan diagram Alur.

A. *Analysis*

A.1 Analisis Proses Bisnis

1. Prosesnya berawal dari pengolahan untuk mendaftarkan koleksi yang akan diolah ke dalam aplikasi pengadaan. Proses mendaftarkan koleksi ke aplikasi pengadaan ini dilakukan oleh kepala bagian perpustakaan.
2. Kemudian penentuan kategori/kelas oleh petugas pengelola pustaka yang mengacu pada buku DDC23 mulai dari pencarian pada kelas utama ke kelas kedua dan kelas ketiga sebagai kelas spesifik atau kelas kecil dari penentuan kategori tersebut. Membutuhkan waktu kurang lebih 3 sampai 10 menit.
3. Dilanjutkan ke pengolahan untuk menentukan data dengan pengindukan dan input deskripsi, Setelah selesai klik simpan maka data buku akan masuk ke sistem *Online Catalog* dan dapat diakses secara umum.
4. Kemudian untuk menuju *Online Catalog* yaitu dengan cara mengakses *Undika Library* lalu memilih *Online Catalog*.
5. Lalu untuk melakukan pencarian koleksi, pemustaka dapat memilih jenis koleksi seperti Buku, Majalah, *Software*, atau Tugas Akhir

6. Selanjutnya memilih kriteria pencariannya seperti Judul, Pengarang, Subyek, atau ISBN dan memasukan kata kunci pencarian.
7. Setelah selesai maka akan tampil hasil dari pencarian koleksi dan untuk mengetahui deskripsi lengkap buku, pemustaka dapat memilih/klik salah satu hasil atau judul yang sesuai.

A.2 Analisis Kebutuhan

A. Kebutuhan Pengguna

Pengelola Pustaka Perpustakaan Undika.

B. Kebutuhan Fungsional

B.1 Melihat data kategori

B.2 Mengelola *Filtering*

B.3 Mengelola Data *Training*.

B.4 *Testing*.

B.5 Mengelola data klasifikasi

C. Kebutuhan Non-Fungsional

C.1 Aksesibilitas

Sistem dapat diakses di manapun dan kapanpun pengguna inginkan secara *online*. Adanya kebutuhan jaringan internet untuk terhubung ke sistem.

C.2 Sistem Berbasis *Website*

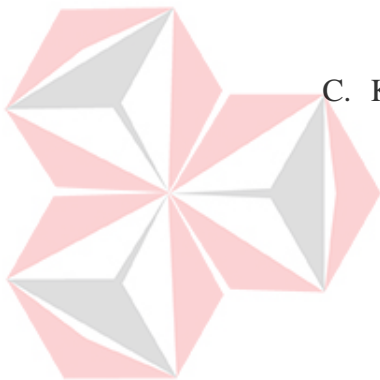
Dapat dijalankan melalui beberapa perangkat lunak *browser* diantaranya adalah *Mozilla Firefox*, *Google Chrome* dan *Internet Explore*.

C.3 Operasional

Perangkat komputer yang digunakan memiliki spesifikasi minimal *Processor Pentium III 500 Hz*, *Memory (RAM) 64 MB*, dan *Harddisk 10 GB*.

C.4 *Usability*

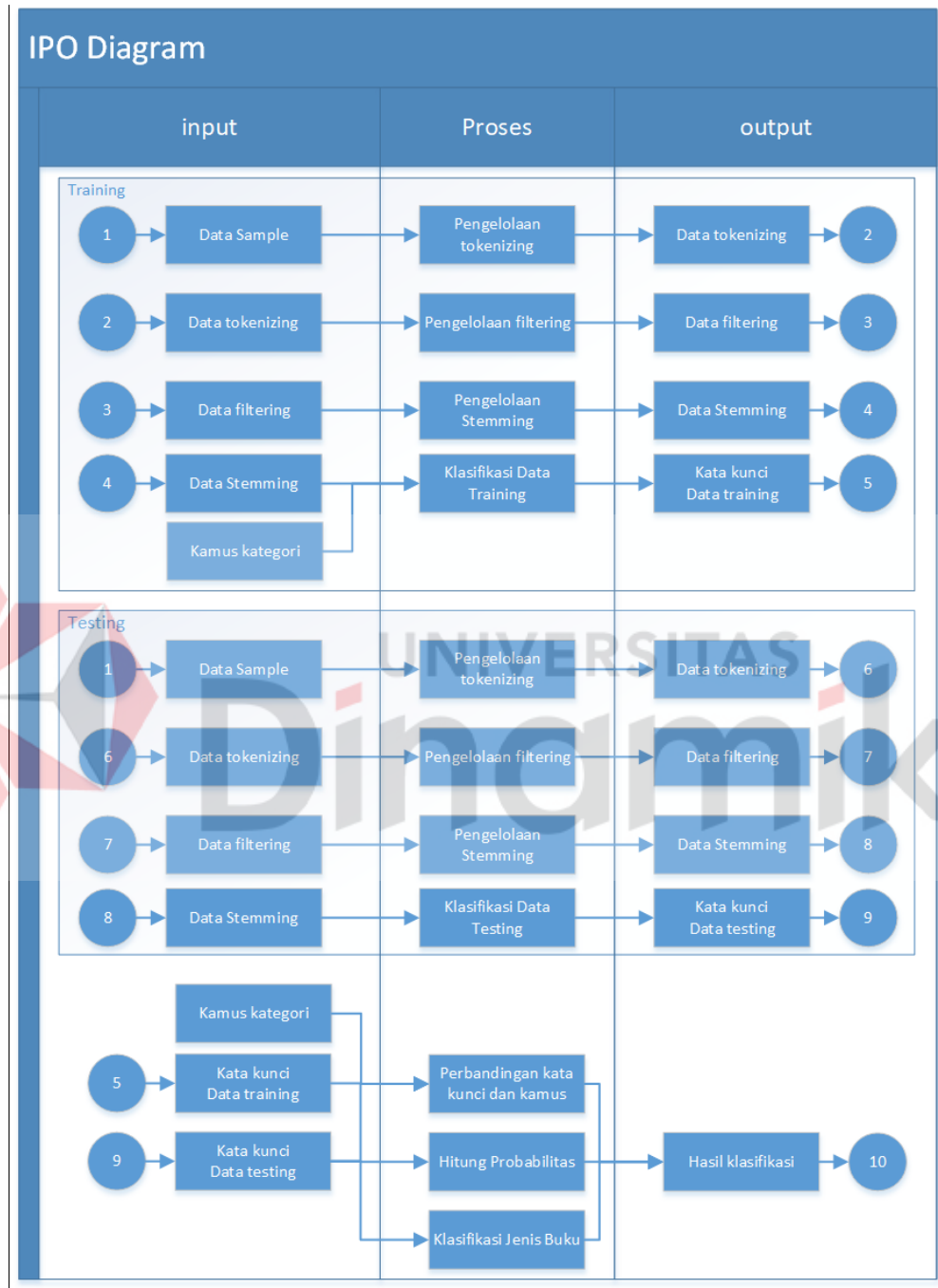
Antarmuka pengguna yang mudah dipahami dan cukup informatif untuk kebutuhan pengguna.



B. Design

B.1 IPO Diagram

Dibawah ini adalah IPO diagram penelitian.



Gambar 3.2 IPO Diagram

Berikut Penjelasan dari Gambar 3.2 IPO diagram diatas :

Input

1. *Data Sample*

Merupakan data yang digunakan untuk implementasi yang terbagi menjadi data *training* dan data *testing*.

2. *Data tokenizing*

Berisi data hasil dari proses *tokenizing* yang mengubah bentuk kalimat ke daftar kata-kata.

3. *Data filtering*

Berisi daftar kata-kata penting yang sudah dicocokkan untuk membuang kata-kata yang tidak diperlukan atau kurang penting menggunakan *Stopword/Stoplist*.

4. *Data stemming*

Berisi daftar kata-kata dasar yang telah diubah dari *filtering* yang akan digunakan untuk kata kunci.

5. *Data Testing*

Sebagai data yang digunakan untuk proses perhitungan klasifikasi buku. Berisi judul atau deskripsi yang akan dimasukan.

6. Kamus Kategori

Berisi kumpulan kata-kata dari kategori tertentu.

Proses

1. *Pengelolaan Tokenizing*

Proses ini dilakukan untuk mengubah bentuk kalimat ke satuan kata-kata.

2. *Pengelolaan filtering*

Bagian ini adalah untuk pengelolaan dari *tokenizing* menjadi kata-kata deskriptif seperti “yang”, “dan”, dan “di” dengan *Stopword* yang digunakan untuk proses stemming.

3. *Pengelolaan Stemming*

Proses ini akan mengubah data dari filtering ke bentuk kata dasar.

4. *Klasifikasi Data Training*

Pada proses ini dilakukan klasifikasi menggunakan kamus kategori.

5. *Klasifikasi Data Testing*



UNIVERSITAS
Dinamika

Pada proses ini dilakukan klasifikasi untuk menghilangkan teks *noise* untuk digunakan pada klasifikasi perbandingan dengan data *training* dan kamus.

6. Perbandingan kata kunci dan kamus

Merupakan proses pemilihan keakuratan kata disesuaikan dengan data buku pengujian dan kamus.

7. Hitung probabilitas

Proses perhitungan probabilitas untuk menentukan klasifikasi.

8. Hitung Klasifikasi

Proses ini yaitu menghitung klasifikasi sesuai rumus NBC untuk menentukan klasifikasi kategori yang sesuai.

Output

1. Data *Tokenizing*

Berisi data hasil dari proses *tokenizing* yang mengubah bentuk kalimat ke daftar kata-kata.

2. Data *Filtering*

Berisi daftar kata-kata penting yang sudah dicocokkan untuk membuang kata-kata yang tidak diperlukan atau kurang penting menggunakan *Stopword/Stoplist*.

3. Data *Stemming*

Berisi daftar kata-kata dasar yang telah diubah dari *filtering* yang akan digunakan untuk kata kunci.

4. Kata kunci data *Training*

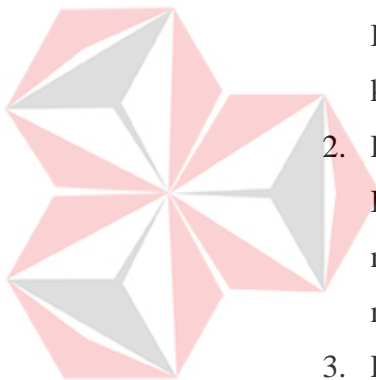
Berisi data kata-kata yang akan dilakukan perbandingan pada tahap *Testing*.

5. Kata kunci data *Testing*

Berisi data kata-kata yang akan dilakukan perbandingan dengan hasil kata kunci pada tahap *Training*.

6. Hasil klasifikasi

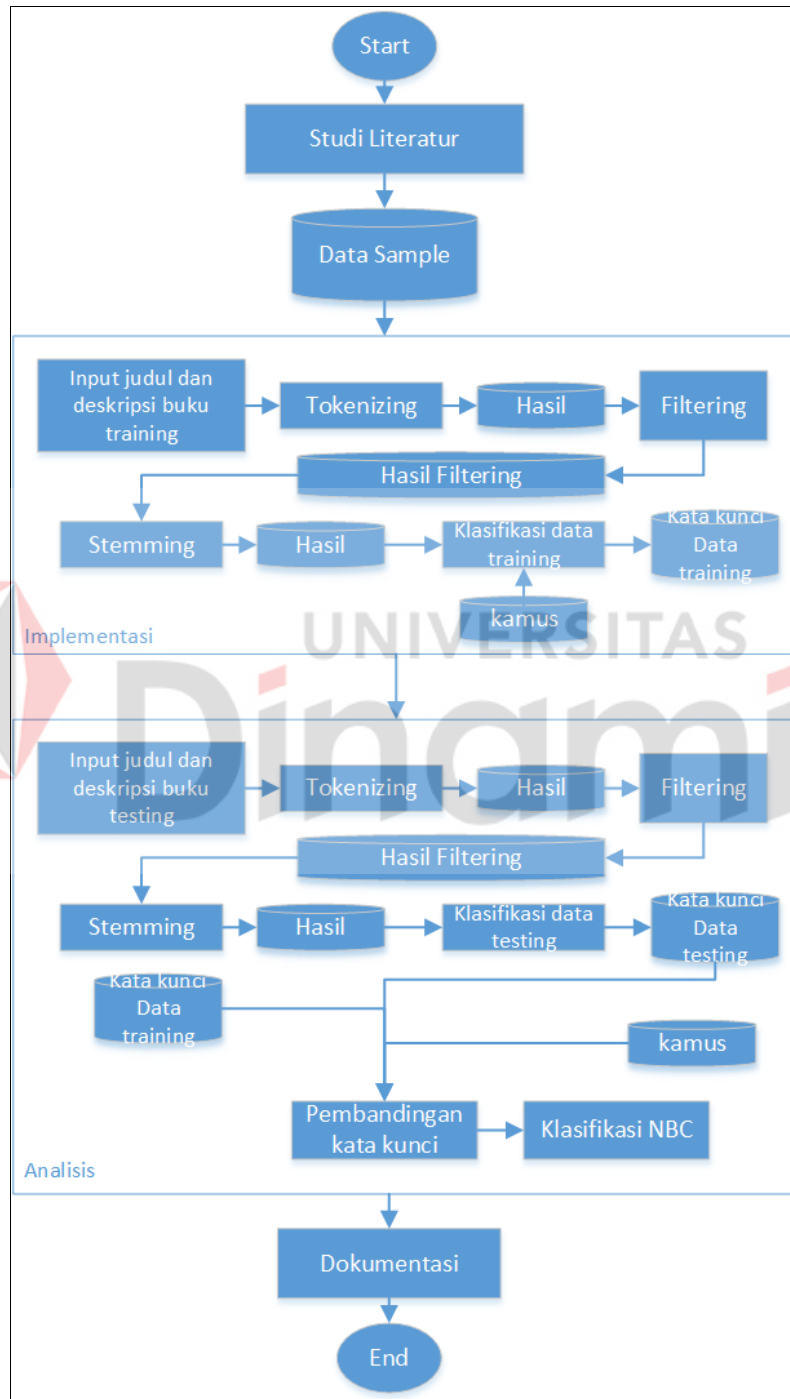
Hasil ini menunjukkan kategori yang sesuai dengan buku yang telah dilakukan *testing* klasifikasi dengan NBC.



Untuk alur lebih jelas tentang pembagian klasifikasi pada dokumen *Training* dan *Testing* ditunjukkan pada gambar 3.3, 3.4 dan 3.5 dibawah.

B.2 Skema Pengembangan Penelitian

Berikut merupakan skema pengembangan pada Penelitian.



Gambar 3.3 Skema Pengembangan

Pada Gambar 3.3 merupakan skema pengembangan yang akan diterapkan pada penelitian ini. Dimulai dari Studi Literatur yaitu peneliti mengumpulkan dan mempelajari jurnal untuk kebutuhan pustaka tentang algoritma dan teori NBC. Proses selanjutnya adalah data internal yang diperlukan yaitu buku-buku koleksi yang berasal dari Perpustakaan Undika yang sudah dilakukan perhitungan dengan rumus Slovin untuk menentukan data sampel sebanyak 389 buku yang terbagi menjadi 311 data training dan 78 data testing untuk uji coba dalam skenario dan mengukur tingkat akurasi NBC. Akurasi hasil klasifikasi ditentukan dengan persamaan berikut :

$$Akurasi = \frac{\text{Jumlah Kategori Benar}}{\text{Jumlah dokumen uji}} \times 100\%$$

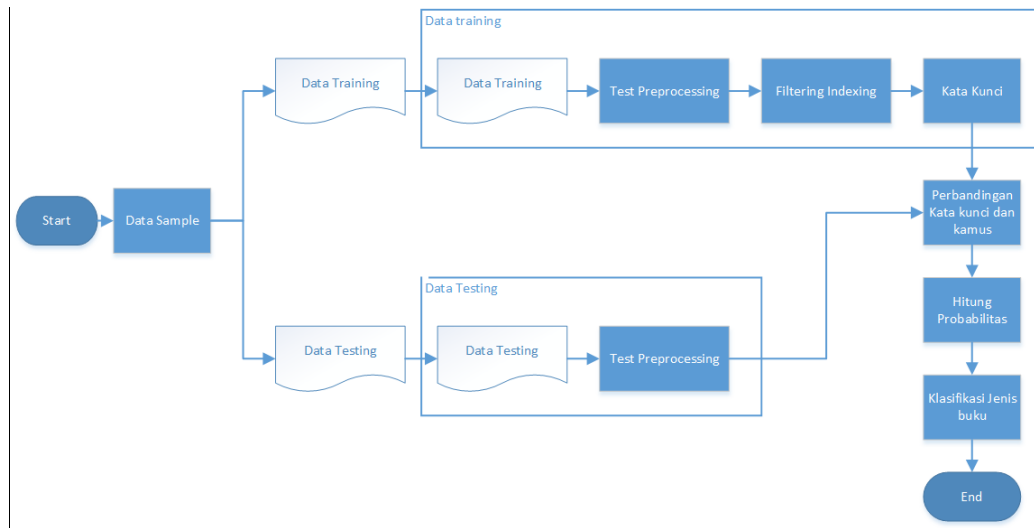
Berikut merupakan rumus Slovin untuk perhitungan data sample:

$$n = \frac{N}{1 + Ne^2}$$

Selanjutnya yaitu proses implementasi mulai dari input judul dan deskripsi buku kemudian dijadikan kata per-kata dengan *Tokenizing* dan hasilnya akan di *Filtering* menggunakan *stopword removal* yang mana menghasilkan kata kata penting untuk diteruskan pada proses *Stemming* menghasilkan reduksi kata ke bentuk dasarnya. Setelah selesai akan di *filter* dan menghasilkan kata kunci. Pada tahap selanjutnya yaitu mendokumentasikan hasil dari keseluruhan prosesnya.

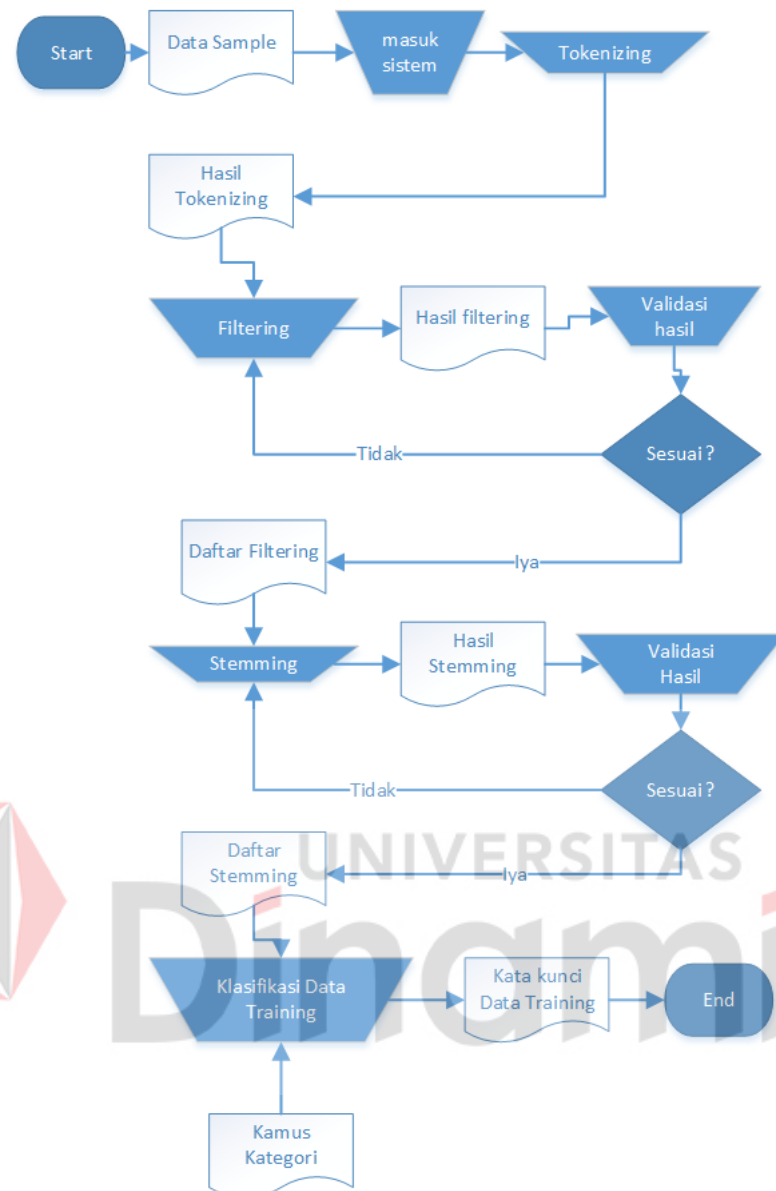
B.3 Diagram Alur Sistem

Berikut merupakan diagram alur klasifikasi NBC.



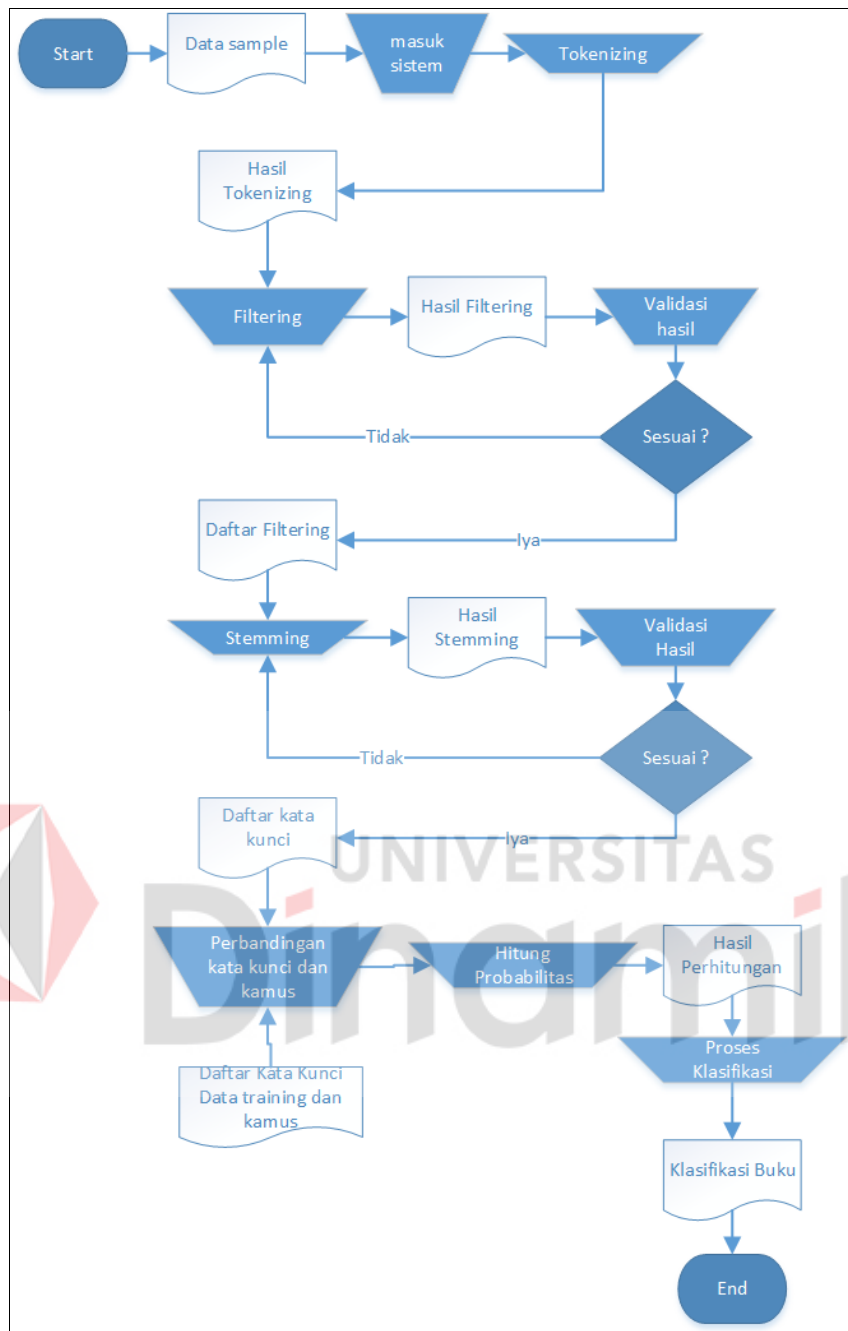
Gambar 3.4 Alur Proses Klasifikasi Naive Bayes

Pada Gambar 3.4 terlihat yaitu dalam metode NBC terdapat 2 tahapan yaitu *training* dan *testing*. Alur detail pada tahap *training* ditunjukkan pada Gambar 8 berikut :



Gambar 3.5 Diagram tahap training

Pada Gambar 3.5 merupakan gambar yang menunjukkan proses *training* yang dilakukan pada data training dari data sample. Proses training diatas dilakukan untuk melakukan klasifikasi data training dengan kamus kategori yang akan menghasilkan data *training*. berikutnya yaitu tahap kalsifikasi yang mana dalam penelitian ini ditunjukkan pada gambar berikut :



Gambar 3. 6 Diagram tahap testing klasifikasi

Pada Gambar 3.6 diatas merupakan tahap uji coba atau *testing* pada data *testing* yang diambil dari data *sample*. Pada proses ini terdapat perbandingan kata kunci dan kamus yang selanjutnya dilakukan perhitungan NBC untuk klasifikasi kategori buku yang sesuai.

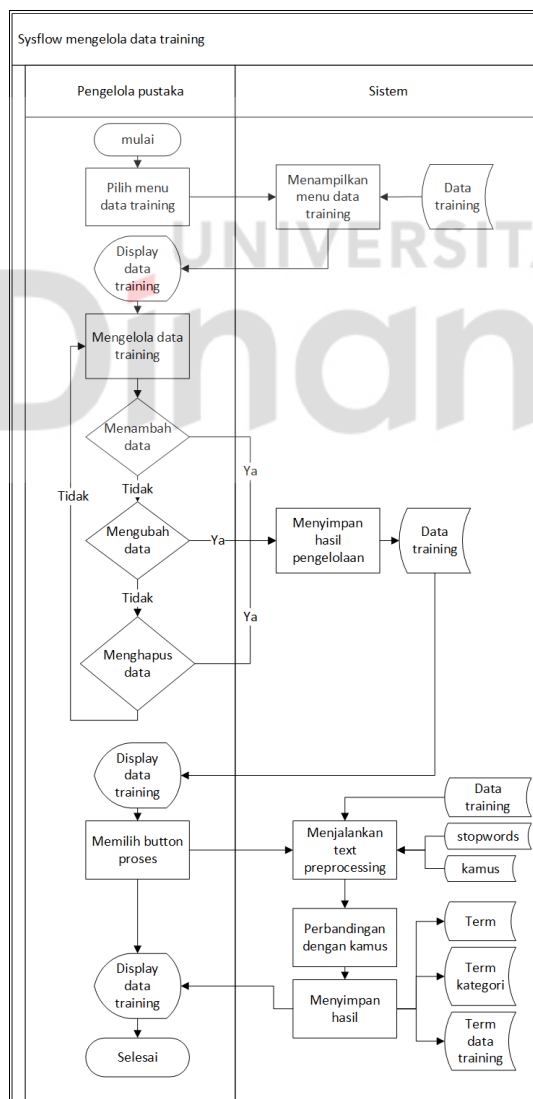
C. Desain Sistem

Berikut ini merupakan rancangan sistem klasifikasi kategori.

C.1 System Flow

Pada *System Flow* ini menggambarkan seluruh proses dari sistem klasifikasi kategori buku. Terdapat 6 proses yang akan ditunjukkan pada gambar – gambar dibawah ini. Untuk desain *System Flow* fungsi pendukung dapat dilihat pada halaman lampiran 2.

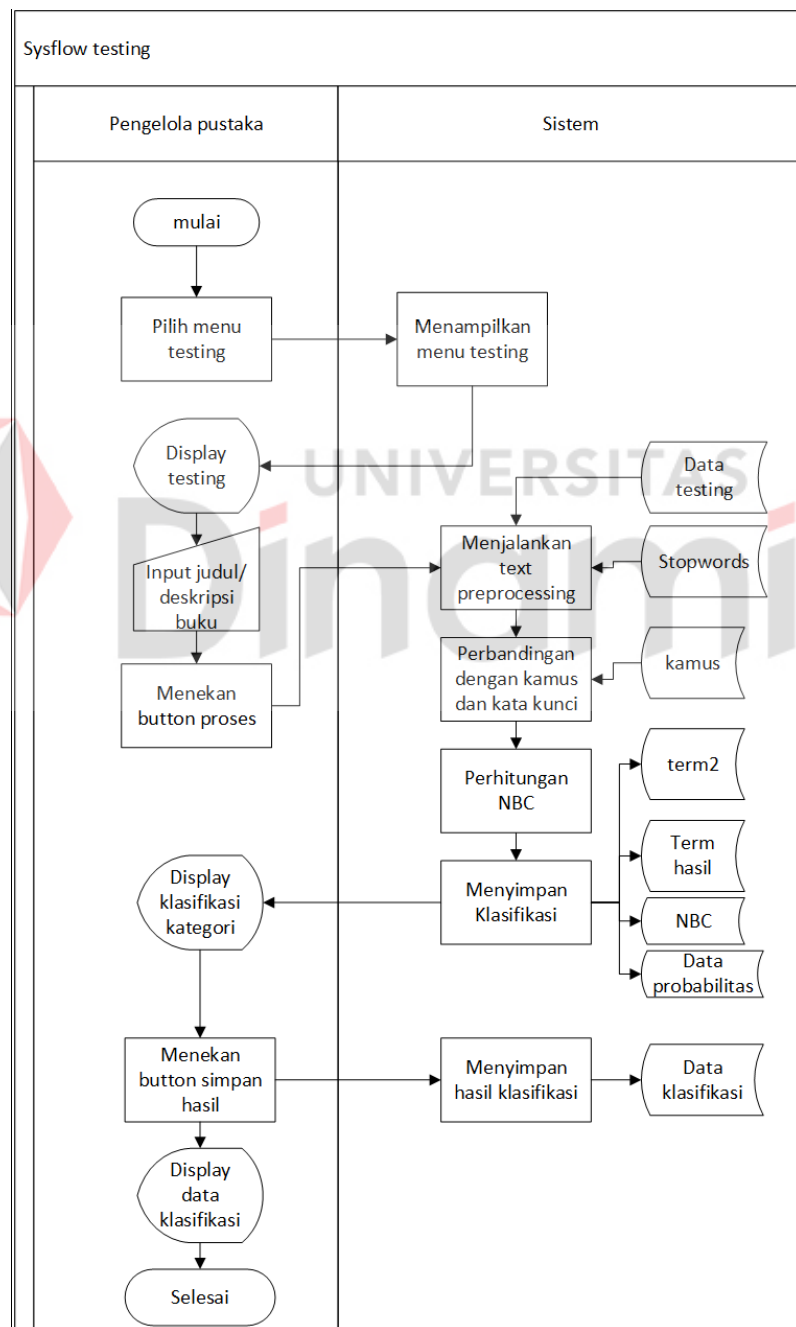
C.1.1 System Flow Mengelola data Training



Gambar 3. 7 System Flow Mengelola data Training

Terlihat pada gambar diatas menggambarkan tentang alur sistem untuk mengelola data *training* dan menjalankan rangkaian proses *text preprocessing* seperti *tokenizing*, *filtering*, dan *stemming* kemudian dilakukan perbandingan dengan kamus kategori yang selanjutnya digunakan untuk proses pengujian.

C.1.2 System Flow Testing



Gambar 3. 8 System Flow Testing

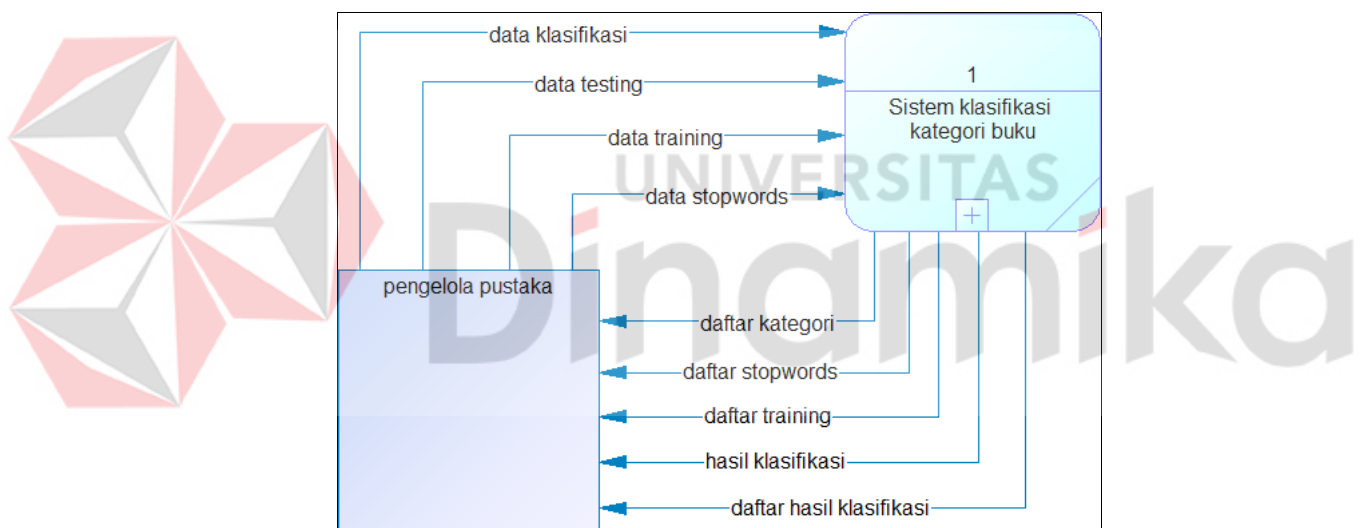
Pada Gambar diatas menunjukkan alur sistem melakukan pengujian data menggunakan rangkaian proses *text preprocessing* yaitu *tokenizing*, *filtering*, dan *stemming* dan menjalankan penghitungan NBC.

C.2 Data Flow Diagram

Dibawah ini merupakan rancangan *Data Flow Diagram* yang telah dibuat.

C.2.1 Context Diagram

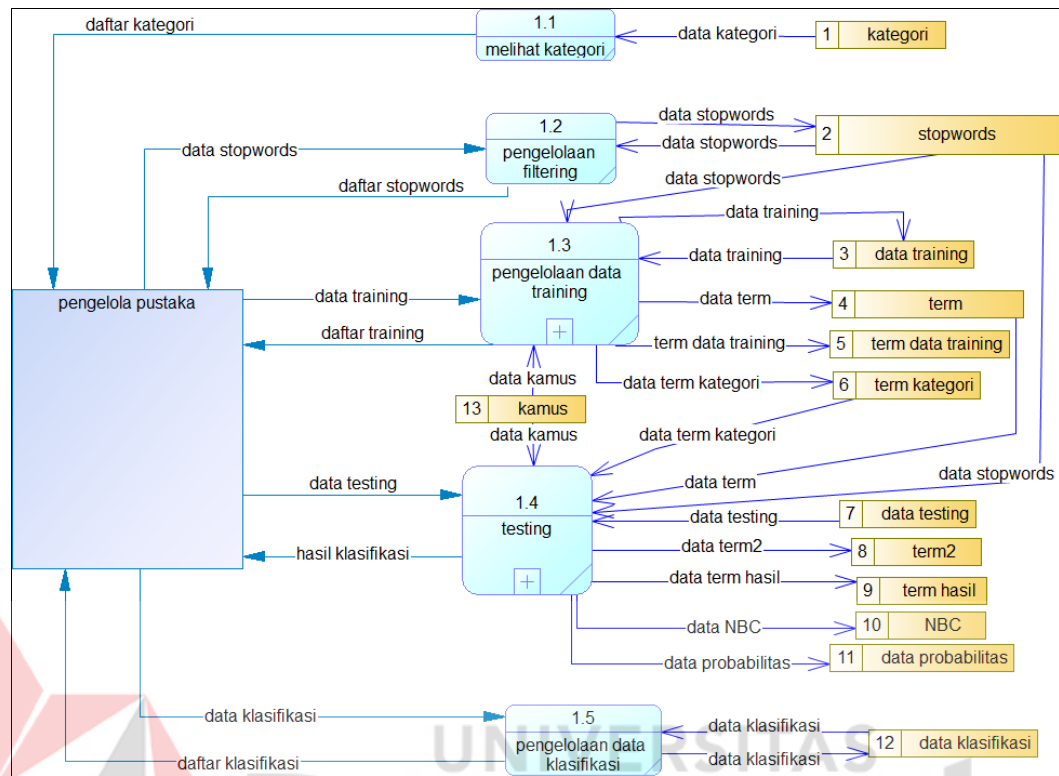
Berikut dibawah ini adalah rancangan *Context Diagram* sistem klasifikasi kategori buku.



Gambar 3. 9 *Context Diagram* sistem

Gambar diatas merupakan *Context Diagram* untuk menunjukkan hubungan entitas dan proses sistem. Pada proses sistem terdapat *decompose process* yang mana prosesnya akan dilanjutkan pada *Data Flow Diagram level 0*.

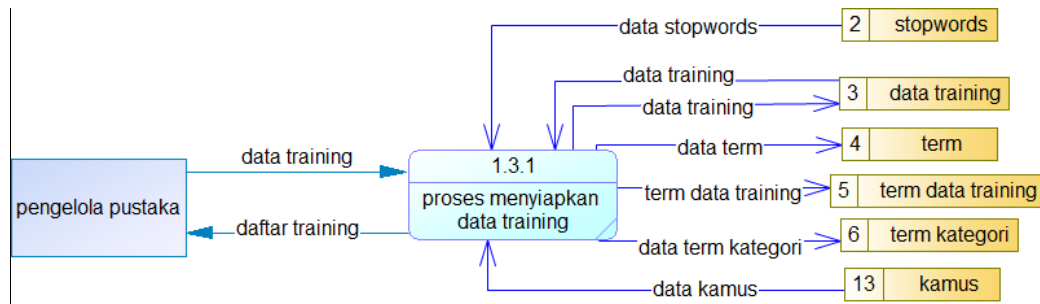
C.2.2 Data Flow Diagram Level 0



Gambar 3. 10 Data Flow Diagram level 0

Gambar diatas menunjukkan hubungan entitas (simbol kotak berwarna biru tua) dengan setiap proses yang ada (simbol kotak berwarna biru muda dengan nomor 1.1 sampai 1.5) pada sistem dengan database (simbol berwarna kuning).

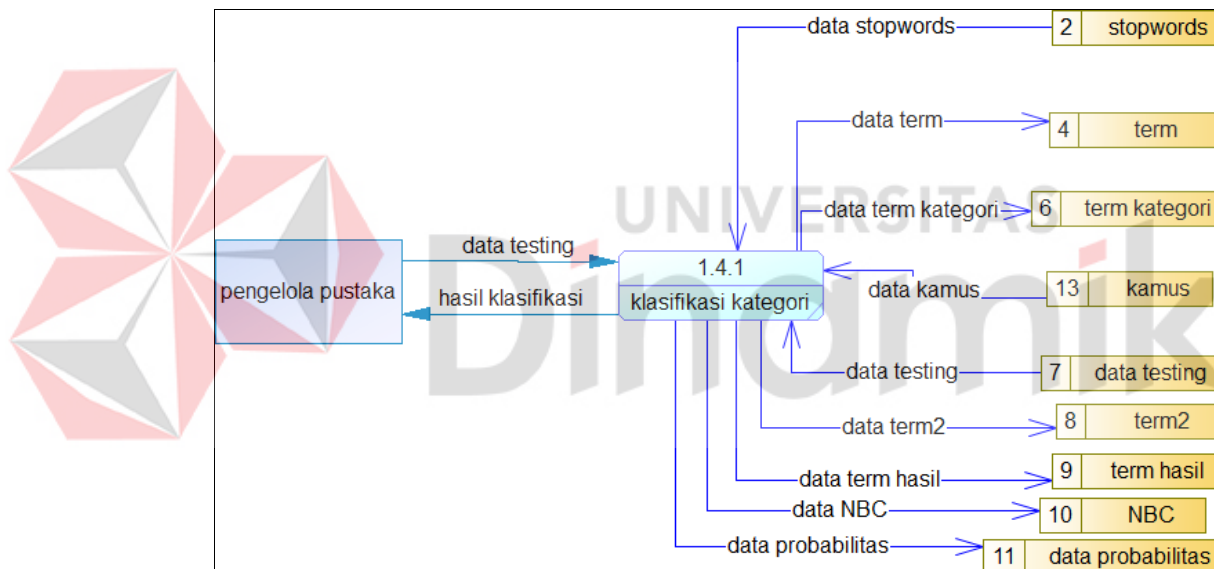
C.2.3 Data Flow Diagram level 1 proses menyiapkan data training



Gambar 3. 11 Data Flow Diagram level 1 proses menyiapkan data training

Pada gambar diatas menunjukan proses *decompose* dari proses pengelolaan data training.

C.2.4 Data Flow Diagram level 1 klasifikasi kategori



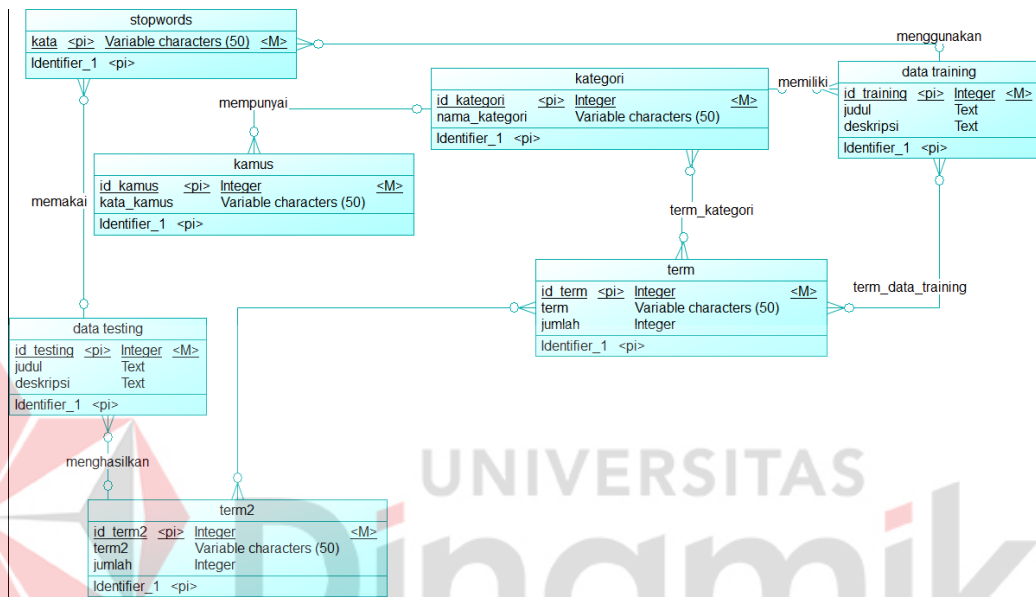
Gambar 3. 12 Data Flow Diagram level 1 klasifikasi kategori

Pada gambar diatas merupakan *decompose* proses dari proses testing pada Data Flow Diagram level 0.

C.3 Merancang Basis Data

Dibawah ini adalah rancangan skema *database* yang akan digunakan sistem. Ditunjukkan dalam bentuk *Conceptual data model*, *physical data model*, dan struktur tabel.

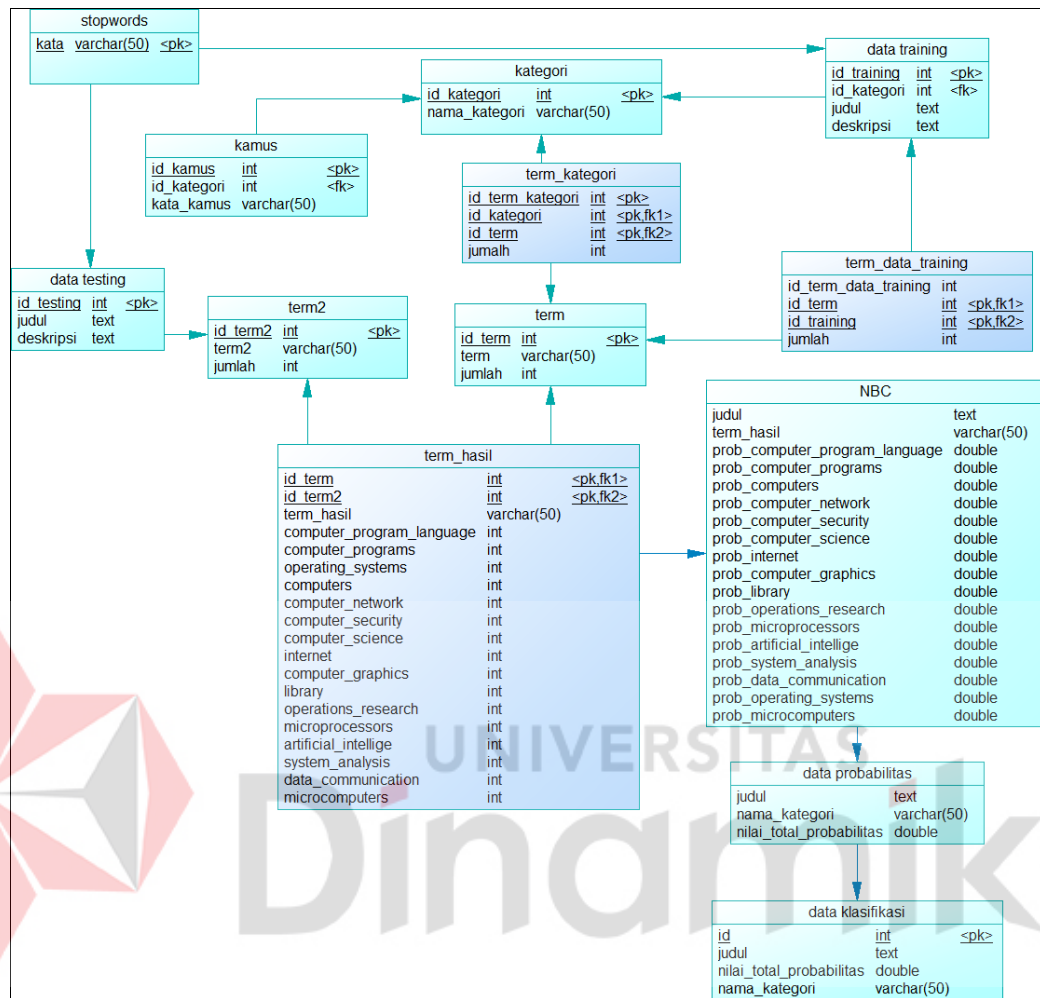
C.3.1 *Conceptual data model* (CDM)



Gambar 3. 13 *Conceptual data model* Sistem

Pada Gambar diatas menunjukan entitas yang dihubungkan dengan entitas lain membentuk relasi. Pada masing masing entitas juga ditunjukkan atribut, tipe data, dan masing - masing panjangnya.

C.3.2 Physical data model (PDM)



Gambar 3. 14 Physical data model Sistem

Dari gambar diatas menunjukan entitas – entitas generate dari CDM. Terdapat entitas baru yang dihasilkan karena relasi yang dibuat berdasarkan *cardinalitas* many to many dan terdapat tabel baru yang dibuat setelah muncul tabel hasil relasi untuk menyimpan kebutuhan data tambahan yaitu NBC, data probabilitas, dan data klasifikasi.

C.4 Struktur tabel

Dibawah ini merupakan struktur tabel berdasarkan PDM dari sistem klasifikasi kategori yang telah dibuat. Struk tabel lebih lengkap dapat dilihat pada lampiran 2.

C.4.1 Tabel kategori

Primary key : id_kategori

Foreign key :

Fungsi : menyimpan data kategori

Tabel 3. 1 Struktur tabel kategori

No.	Field	Data type	length	Constraint
1.	id_kategori	int		<i>Primary key</i>
2.	nama_kategori	varchar(50)	50	

C.4.2 Tabel data *Training*

Primary key : id_training

Foreign key :id_kategori

Fungsi : menyimpan data *Training*

Tabel 3. 2 Struktur tabel *Training*

No.	Field	Data type	length	Constraint
1.	id_training	int		<i>Primary key</i>
2.	id_kategori	int		<i>Foreign key</i>
3.	judul	text		
4.	deskripsi	text		

C.4.3 Tabel data testing

Primary key : id_testing

Foreign key :

Fungsi : menyimpan data testing

Tabel 3. 3 Struktur tabel data testing

No.	Field	Data type	length	Constraint
1.	id_testing	int		<i>Primary key</i>
2.	judul	text		
3.	deskripsi	text		

C.5 Desain *Interface*

Dibawah ini merupakan gambaran *interface* pada sistem klasifikasi kategori. *Interface* untuk fungsi pendukung dapat dilihat pada halaman lampiran.

C.5.1 *Interface Data Training*

DATA TRAINING

ID TRAINING
 JUDUL
 DESKRIPSI
 ID KATEGORI

TAMBAH ➔

Pilih berkas Choose File No file chosen Import

Delete All data

ID Training	Judul	Deskripsi	ID Kategori	Option
id training 001	txt judul	txt deskripsi	id kategori 001	Ubah hapus

PROSES ➔

Gambar 3. 15 *Interface data training*

Gambar diatas merupakan tampilan untuk memasukan data training. Dengan desain konsisten memakai *background* kontras kombinasi warna ungu dan biru. Daftar data training ditampilkan rapih berbentuk tabel dibuat transparan dengan button berwarna biru.

C.5.2 Interface Testing

Gambar 3. 16 Interface Testing

Pada Gambar diatas menunjukan form input yang di desain menarik mata pengguna, font italic dan *background* transparan blur dengan gambar seseorang memegang buku diantara kumpulan buku.

C.5.3 Interface Klasifikasi Kategori

Peringkat	Judul	Nilai probabilitas	Kategori
1	txt judul	123	txt nama kategori
2	txt judul	123	txt nama kategori
3	txt judul	123	txt nama kategori

Gambar 3. 17 Interface klasifikasi kategori

Tampilan yang ditunjukkan pada Gambar diatas mempunyai desain yang sama dengan halaman sebelumnya agar terlihat serasi dan kombinasi yang bagus dari warna tabel putih dan warna background gradient biru dan ungu.

HASIL KLASIFIKASI DAN PERHITUNGAN NBC

T a b e l 1 Jumlah kata kunci

Judul	Term Hasil	computer program language (005)	computer programs (005)	operating systems (005)	computers (005)	computer network (004)	computer security (005)	computer science (005)	internet (004)	computer graphics (006)	library (020)	operations research (003)	microprocessors (004)	artificial intelligence (006)	system analysis (004)	data communication (004)	microcomputers (004)
tit judul	tit term	84	312	96	29	46	33	24	50	21	4	2	2	50	24	0	0

T a b e l 2 Hasil perhitungan probabilitas

Judul	Term Hasil	computer program language(005)	computer programs(005)	operating systems(005)	computers(005)	computer network(004)	computer security(005)	computer science(005)	internet(004)
tit judul	tit term	0.029772320246935	0.067529665587918	0.058113636363636	0.031120331930207	0.015068932350112	0.022880215343203	0.025125628140704	0.037253469689022

Judul	Term Hasil	computer graphics(006)	library(020)	operations research(003)	microprocessors(004)	artificial intelligence(006)	system analysis(004)	data communication(004)	microcomputers(004)
tit judul	tit term	0.014814814814815	0.0063371356147022	0.0034013605442177	0.0040927694406548	0.052307692307692	0.034626038781163	0.001980198019802	0.002247191011236

T a b e l 3 Nilai masing-masing kategori

Peringkat	Judul	Nilai probabilitas	Kategori
1	tit judul	123	tit nama kategori
2	tit judul	123	tit nama kategori
3	tit judul	123	tit nama kategori
4	tit judul	123	tit nama kategori

Gambar 3. 18 Interface detil klasifikasi kategori

Gambar diatas menunjukkan tampilan detil hasil klasifikasi dan perhitungan NBC dari masing masing kategori.

3.1.4 Construction

Pada tahap ini, akan mulai merancang aplikasi atau pengkodean berdasarkan desain yang telah dibuat. Pengembangan aplikasi ini dilakukan dari awal hingga aplikasi siap dijalankan.

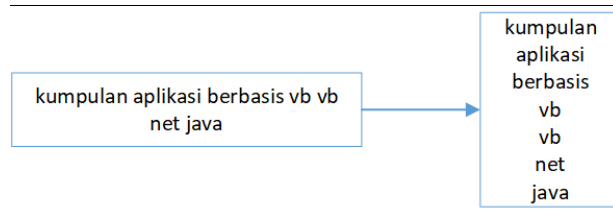
Tahapan awal dari klasifikasi teks ialah *Text Teks preprocessing* yang bertujuan untuk mempersiapkan teks menjadi data yang akan mengalami pengolahan pada tahapan berikutnya.

A.1 *Case Folding* yaitu mengubah semua huruf dalam teks data *Training* menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter. Kemudian dilanjutkan ke tahap selanjutnya yaitu *Tokenizing*. Untuk membuat judul dan deskripsi buku hanya huruf a sampai z menggunakan fungsi *strtolower*. Berikut ini adalah penerapannya.



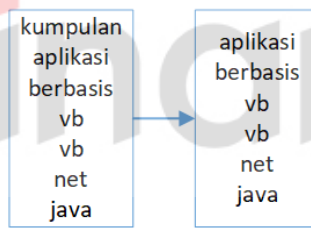
Gambar 3. 19 proses Case folding

A.2 *Tokenizing* adalah sebuah proses untuk memilah isi teks sehingga menjadi satuan kata-kata. Proses ini mengambil data dari hasil *Case Folding* untuk membuat kata per kata menggunakan fungsi *str_replace*. Berikut ini adalah penerapannya.



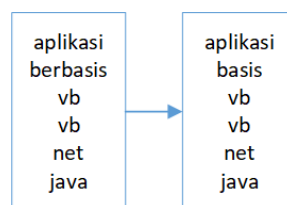
Gambar 3. 20 proses *Tokenizing*

A.3 Tahapan *Filtering* merupakan tahap mengambil kata-kata penting dari hasil *token*. menggunakan algoritma *Stoplist* (membuang kata yang kurang penting) atau *wordlist* (menyimpan kata penting). Menggunakan perulangan *for* dan pengkondisian *if else* untuk memilih hanya kata – kata yang selain dari data *filtering* yang sudah disiapkan menggunakan tabel. Berikut ini adalah penerapannya.



Gambar 3. 21 proses *filtering*

A.4 *Stemming* merupakan suatu proses untuk mereduksi kata ke bentuk dasarnya untuk menghilangkan imbuhan – imbuhan yang ada pada setiap kata menggunakan fungsi *preg_replace*. Berikut ini adalah penerapannya.



Gambar 3. 22 proses *Stemming*

Kemudian dilanjutkan dengan proses perbandingan dengan kamus kategori untuk menghasilkan kata kunci sebagai hasil dari klasifikasi data *Training* dan digunakan untuk *Testing*.

3.1.5 Deployment

Tahap akhir akan dilakukan pengujian. Untuk mengetahui fungsi – fungsi yang ada telah sesuai dengan kebutuhan. Berikut ini adalah fungsi – fungsi yang diuji coba.

A. Skenario Pengujian

Pada Sistem klasifikasi kategori ini terdapat dua pengujian yaitu *Black Box* untuk menguji fungsional sistem dan pengujian akurasi.

A.1 Pengujian *Black Box*

Berikut ini merupakan daftar fungsi – fungsi yang akan diuji lengkap dengan *Test Case* yang digunakan. Untuk fungsi pendukung dapat dilihat pada halaman lampiran.

Tabel 3. 4 *Test Case*

No	Fungsi	Test Case
1	Melihat data kategori	TC0101
2	Mengelola <i>Filtering</i> tambah data	TC0201
		TC0202
		TC0203
3	Mengelola <i>Filtering</i> hapus data	TC0301
4	Mengelola data <i>Training</i> tambah data	TC0401
		TC0402
		TC0403
5	Mengelola data <i>Training</i> hapus data	TC0501
6	Mengelola data <i>Training</i> ubah data	TC0601
		TC0602
		TC0603
7	Proses menyiapkan data <i>Training</i>	TC0701
		TC0702
		TC0703
8	<i>Testing</i>	TC0801
		TC0802
		TC0803
9	Mengelola data klasifikasi	TC0901
		TC0902

A.1.1 Proses menyiapkan data *Training*

Tabel 3. 5 hasil pengujian proses menyiapkan data *Training*

<i>Test Case</i>	<i>Expected Result</i>	<i>Result</i>
TC0701	Berhasil, dan muncul <i>pop up</i> klasifikasi data training selesai.	Berhasil
TC0702	Gagal, data Training kosong.	
TC0703	Berhasil, dan kembali ke halaman <i>Training</i>	Berhasil

A.1.2 *Testing*

Tabel 3. 6 hasil pengujian *Testing*

<i>Test Case</i>	<i>Expected Result</i>	<i>Result</i>
TC0801	Berhasil, dan muncul <i>pop up</i> klasifikasi data training selesai.	Berhasil
TC0802	Gagal, inputan testing tidak terisi.	Berhasil
TC0803	Berhasil, dan menampilkan halaman klasifikasi menunjukkan hasil klasifikasi kategori dari data <i>Testing</i> .	Berhasil

A.2 Pengujian Akurasi

Pengujian Akurasi klasifikasi kategori dilakukan dengan membandingkan klasifikasi kategori dari Perpustakaan Undika dan klasifikasi kategori yang dilakukan menggunakan sistem dengan algoritma NBC. Pengujian menggunakan 311 data *Training* dan 78 data *Testing*. Hasil klasifikasi oleh Perpustakaan Undika dapat dilihat pada tabel dibawah ini. Hasil klasifikasi lengkap oleh Perpustakaan Undika dapat dilihat dalam tabel pada halaman lampiran.

Tabel 3. 7 hasil klasifikasi kategori Perpustakaan Undika

No	Judul	Deskripsi	Klasifikasi oleh Perpustakaan Undika
1	MEMBUAT APLIKASI DATABASE DENGAN <i>java</i> 2: PANDUAN APLIKATIF	Buku ini membahas tentang teknik pembuatan aplikasi <i>database</i> melalui <i>java</i> 2. Pembahasan meliputi pengenalan tentang <i>database</i> , SQL, persiapan lingkungan pemrograman <i>java</i> 2, mendesain modul validasi input data, proses data, dan autentifikasi <i>user</i> , hingga contoh-contoh pembuatan aplikasi Sistem Informasi Data, yaitu Penjualan dan Perpustakaan. Diharapkan setelah mempelajari buku ini pembaca dapat menguasai dan mampu membuat aplikasi <i>database</i> berbasis <i>java</i> 2 sesuai kebutuhan.	<i>Computer program language</i>
28	SISTEM OPERASI MICROSOFT WINDOWS 2000 PROFESSIONAL : BELAJAR SENDIRI, EDISI BARU	Sistem Operasi <i>Windows</i> 2000 adalah suatu sistem operasi yang dibangun diatas Teknologi NT (<i>Network Teknologi</i>). Lebih tepatnya gabungan <i>Windows</i> 98 dengan <i>Windows</i> NT. Karena <i>Microsoft</i> sudah memandang jauh kedepan bahwa teknologi masa mendatang adalah teknologi <i>network</i> . Seperti internet, LAN (<i>Local Area Network</i>) dan lain-lain.	<i>computer programs</i>
33	MANDRIVA LINUX, LIMITED EDITION 2005	Buku ini sangat berguna bagi anda yang menggunakan <i>linux</i> sebagai sistem operasi komputer untuk pekerjaan sehari-hari dikantor, warnet pendidikan dan di rumah. Dalam buku ini semua aplikasi penting dibahas secara praktis dan sistematis. Buku ini terbagi beberapa Bab, mulai dari jalur cepat mempelajari <i>mandriva</i> , petunjuk menginstalasi termasuk cara mempartisi hard disk, konfigurasi piranti keras mulai dari mengkonfigurasi <i>VGA Card</i> , <i>moniotr</i> , <i>sound card</i> , <i>printer</i> , kartu jaringan dan juga modem. Selain itu membahas tentang jaringan dan internet seperti <i>web</i> , <i>e-mail</i> , <i>HTML editor</i> , file sharing, FTP dan juga remote login. Anda akan dikenalkan juga pada multimedia dan game.	<i>Operating systems</i>
34	PROSIDING SEMINAR NASIONAL ILMU KOMPUTER DAN APLIKASINYA (SNIKA) 2006, VOLUME 1, NO.1, NOVEMBER 2006	SNIKA 2006 ini merupakan forum nasional pertukaran informasi dan pengetahuan antara pihak akademisi dan praktisi atau industri. Kegiatan ini bertujuan untuk memperlihatkan peranan ilmu komputer sebagai ilmu dasar dalam perkembangan teknologi, khususnya teknologi informasi dan komunikasi. Harapannya acara ini benar - benar dapat menjembatani antara kebutuhan industri dengan penelitian di pihak perguruan tinggi, sehingga membantu kemajuan ilmu dan teknologi komputer dan komunikasi di Indonesia.	<i>Computers</i>

No	Judul	Deskripsi	Klasifikasi oleh Perpustakaan Undika
37	KAMUS LENGKAP JARINGAN KOMPUTER	Bagi orang awam atau orang yang ingin mempelajari komputer pada umumnya dan jaringan komputer pada khususnya, pemahaman terhadap istilah-istilah dalam bidang jaringan komputer akan membantu dalam memahami artikel atau bacaan tentang jaringan komputer. Buku ini disusun dengan tujuan untuk membantu memahami berbagai istilah jaringan komputer dan dapat menjadi acuan untuk memahami istilah-istilah tersebut sehingga proses pemahaman dapat menjadi lebih baik.	<i>computer network</i>



UNIVERSITAS
Dinamika

BAB IV

HASIL DAN PEMBAHASAN

4.1 Implementasi Sistem

Tahap implementasi sistem akan memberikan penjelasan dari rancangan sistem yang sudah dibuat untuk menunjukkan bagaimana cara kerja Sistem Klasifikasi Kategori pada Perpustakaan Undika. Fungsi pendukung lainnya dapat dilihat pada halaman lampiran 3.

4.1.1 Mengelola data Training

ID Training	Judul	Deskripsi	ID Kategori	Option
1	LOGIKA PEMROGRAMAN MENGGUNAKAN C++	Buku ini sangat cocok digunakan untuk pelajar, mahasiswa, atau siapa saja yang bermaksud untuk mempelajari pemrograman komputer menggunakan Bahasa C++. Buku ini lebih menekankan pada cara untuk menyelesaikan masalah. Oleh karena itu, banyak contoh permasalahan yang diberikan dan cara untuk menyelesaikannya. Contoh contoh yang cukup banyak dan bahasa yang mudah dipahami membuat buku ini sangat mudah digunakan dan dapat menjadi penuntun untuk mempelajari Bahasa C++ secara mandiri.	1	<button>Ubah</button> <button>hapus</button>

Gambar 4.1 Halaman Data *Training*

Halaman pada Gambar diatas merupakan untuk mengelola data *Training* yaitu tambah, ubah, hapus, dan menampilkan data *Training*. *Button process* yang berada di bagian bawah merupakan *Button* untuk melakukan klasifikasi data *Training*.

4.1.2 Testing

Gambar 4.2 halaman *Testing*

Gambar diatas merupakan halaman input data testing berisi judul dan deskripsi buku yang akan dilakukan *Testing* klasifikasi kategori, data akan diproses setelah pengguna yaitu pengelola pustaka menekan *Button* proses.

4.1.3 Klasifikasi Kategori

Peringkat	Judul	Nilai probabilitas	Kategori
1	belajar excel untuk perkantoran	0.10938511326861	computer programs(005)
2	belajar excel untuk perkantoran	0.034170854271357	computer science(005)
3	belajar excel untuk perkantoran	0.0092307692307692	artificial intelligence(006)

Gambar 4. 3 halaman klasifikasi Kategori

Pada Gambar diatas menunjukan halaman hasil klasifikasi kategori yang diproses oleh sistem sesuai tahap pertahap yang sudah dirancang dan klasifikasi menggunakan algoritma NBC.

HASIL KLASIFIKASI DAN PERHITUNGAN NBC																		
T a b e l 1 Jumlah kata kunci																		
Judul	Term Hasil	computer program language (005)	computer programs (005)	operating systems (005)	computers (005)	computer network (004)	computer security (005)	computer science (005)	Internet (004)	computer graphics (006)	library (020)	operations research (003)	microprocessors (004)	artificial intelligence (006)	system analysis (004)	data communication (004)	microcomputers (004)	
belajar excel untuk perkantoran	excel	0	506	0	3	0	0	33	0	0	0	0	0	6	0	0	0	
T a b e l 2 Hasil perhitungan probabilitas																		
Judul	Term Hasil	computer program language(005)	computer programs(005)	operating systems(005)	computers(005)	computer network(004)	computer security(005)	computer science(005)	Internet(004)									
belajar excel untuk perkantoran	excel	0.00035026269702277	0.10938511326861	0.00056818181818182	0.004149377593361	0.00032061558191728	0.00067294751009421	0.034170854271357	0.000730460189919									
Judul	Term Hasil	computer graphics(006)	library(020)	operations research(003)	microprocessors(004)	artificial intelligence(006)	system analysis(004)	data communication(004)	microcomputers(004)									
belajar excel untuk perkantoran	excel	0.00067345067340667	0.0012674271229404	0.0011337868480726	0.0013642564802183	0.0092307692307692	0.0013850415612465	0.001980198019802	0.00224719101123									
T a b e l 3 Nilai masing-masing kategori																		
Peringkat	Judul	Nilai probabilitas							Kategori									
1	belajar excel untuk perkantoran	0.10938511326861							computer programs(005)									
2	belajar excel untuk perkantoran	0.034170854271357							computer science(005)									

Gambar 3. 23 halaman detail klasifikasi kategori

Pada gambar diatas menunjukan halaman detail hasil dari klasifikasi dan perhitungan NBC dari masing – masing kategori.

4.2 Pengujian Sistem

Dibawah ini merupakan pengujian sistem yang dilakukan untuk mengetahui fungsi – fungsi yang ada telah sesuai dengan kebutuhan dan sesuai fungsinya. Hasil pengujian lebih lengkap ada pada halaman lampiran 3.

4.2.1. Proses klasifikasi data *Training*

Tabel 4.1 hasil pengujian proses klasifikasi data *Training*

<i>Test Case</i>	<i>Expected Result</i>	<i>Result</i>
TC0701	Berhasil, dan muncul <i>pop up</i> klasifikasi data training selesai.	Berhasil
TC0702	Gagal, data Training kosong.	
TC0703	Berhasil, dan kembali ke halaman <i>Training</i>	Berhasil

4.2.2. *Testing*

Tabel 4.2 hasil pengujian *Testing*

<i>Test Case</i>	<i>Expected Result</i>	<i>Result</i>
TC0801	Berhasil, dan muncul <i>pop up</i> klasifikasi data training selesai.	Berhasil
TC0802	Gagal, inputan testing tidak terisi.	Berhasil
TC0803	Berhasil, dan menampilkan halaman klasifikasi menunjukan hasil klasifikasi kategori dari data <i>Testing</i> .	Berhasil

4.3 Hasil Pengujian

Tabel 4. 3 Hasil pengujian

No	Judul	Klasifikasi oleh Perpustakaan Undika	Kategori sistem
1	membuat aplikasi <i>database</i> dengan <i>java</i> 2: panduan <i>aplikatif</i>	<i>Computer program language</i>	<i>computer program language</i>
2	membuat laporan pdf berbasis web dengan php 5.0	<i>Computer program language</i>	<i>computer program language</i>
3	pemrograman c dan c++	<i>Computer program language</i>	<i>computer program language</i>
4	tuntunan pemrograman <i>java</i> , jilid 3	<i>Computer program language</i>	<i>data communication</i>
5	pemrograman client <i>server microsoft visual basic</i> 6.0, jilid 2	<i>Computer program language</i>	<i>computer program language</i>
6	pengolahan <i>database mysql</i> 5 dengan <i>java</i> 2 disertai teknik pencetakan laporan	<i>Computer program language</i>	<i>computer program language</i>
7	pengembangan aplikasi sistem informasi akademik berbasis sms dengan <i>java</i>	<i>Computer program language</i>	<i>computer program language</i>
8	bank soal <i>visual basic</i>	<i>Computer program language</i>	<i>Computer program language</i>
9	petunjuk praktis membangun aplikasi php <i>point of sale</i>	<i>Computer program language</i>	<i>computer program language</i>
10	web <i>database</i> menggunakan <i>mysql</i> dan <i>sqlite</i>	<i>Computer program language</i>	<i>computer programs</i>
11	<i>visual basic</i> : membuat animasi dan tampilan cantik pada <i>interface form</i>	<i>Computer program language</i>	<i>computer programs</i>
12	pemrograman <i>visual</i> dengan c++ <i>builder</i>	<i>Computer program language</i>	<i>computer program language</i>
13	aplikasi <i>excel</i> dalam mengolah data transaksi dan laporan penjualan	<i>computer programs</i>	<i>computer programs</i>
14	cara mudah menggunakan <i>spss data entry</i>	<i>computer programs</i>	<i>computer programs</i>
15	aplikasi <i>excel</i> dalam <i>pivottable</i> bisnis terapan	<i>computer programs</i>	<i>computer programs</i>
16	panduan lengkap <i>spss</i> 13.0 dalam mengolah data statistik	<i>computer programs</i>	<i>computer programs</i>
17	desain aplikasi penggajian dan pph pasal 21 menggunakan <i>microsoft access</i>	<i>computer programs</i>	<i>computer programs</i>
18	aplikasi <i>excel</i> untuk analisis keuangan	<i>computer programs</i>	<i>computer programs</i>
19	cara mudah menggunakan <i>crystal reports xi</i> : pembuat laporan dari berbagai format data lokal dan <i>sever</i>	<i>computer programs</i>	<i>computer program language</i>
20	strategi jitu memilih metode statistik penelitian dengan <i>spss</i>	<i>computer programs</i>	<i>computer programs</i>
21	utility <i>windows password recovery</i>	<i>computer programs</i>	<i>computer security</i>
22	<i>microsoft visual foxpro</i> 9.0: seri panduan lengkap	<i>computer programs</i>	<i>computer programs</i>
23	<i>microsoft excel</i> 2003, edisi kedua: buku materi kuliah stikom	<i>computer programs</i>	<i>computer programs</i>
24	<i>microsoft acces</i> 2003	<i>computer programs</i>	<i>computer programs</i>
25	7 jam belajar <i>autocad</i> 2006 untuk orang awam	<i>computer programs</i>	<i>artificial intellige</i>
26	mengolah data spasial dengan <i>mapinfo professional</i>	<i>computer programs</i>	<i>computer programs</i>
27	<i>microsoft excel</i> 2000 for <i>windows</i> 98/2000 : belajar sendiri, edisi baru	<i>computer programs</i>	<i>computer programs</i>

No	Judul	Klasifikasi oleh Perpustakaan Undika	Kategori sistem
28	sistem operasi <i>microsoft windows 2000 professional</i> : belajar sendiri, edisi baru	computer programs	computer network
29	buku mini info <i>linux redhat 9</i>	Operating systems	operating systems
30	<i>linux</i> : singkat tepat jelas	Operating systems	computer network
31	sistem operasi, edisi ke2	Operating systems	operating systems
32	panduan belajar sistem operasi	Operating systems	operating systems
33	<i>mandriva linux, limited edition 2005</i>	Operating systems	operating systems
34	<i>prosiding seminar nasional ilmu komputer dan aplikasinya (snika) 2006, volume 1, no.1, november 2006</i>	Computers	computers
35	<i>pc troubleshooting plus</i> , edisi revisi kedua	Computers	Computers
36	18 program jitu untuk meningkatkan kinerja komputer	Computers	computer science
37	kamus lengkap jaringan komputer	computer network	computer network
38	<i>multihoming</i> menggunakan bgp (<i>border gateway protocol</i>) : membangun multi-koneksi ke multi-isp	computer network	computer network
39	jaringan komputer : buku materi kuliah stikom, edisi pertama	computer network	computer network
40	pengelolaan sistem jaringan : modul praktikum	computer network	computer network
41	membangun warnet dan intranet dengan teknologi diskless thin-client linux terminal <i>server project</i> versi 3	computer network	operating systems
42	praktikum <i>microsoft windows server 2003</i> mengkonfigurasi jaringan dan <i>web server</i>	computer network	computer network
43	computer networking first step	computer network	computer network
44	konsep jaringan komputer dan pengembangannya	computer network	computer network
45	panduan membangun jaringan komputer untuk pemula	computer network	computer network
46	meningkatkan kemampuan jaringan komputer dengan <i>pc cloning system</i>	computer network	computer network
47	kamus++ jaringan komputer	computer network	computer network
48	jaringan komputer	computer network	computer network
49	tehnik proteksi : <i>prefentif</i> pribadi sebelum penetrasi terjadi	computer security	computer security
50	kriptografi keamanan internet dan jaringan komunikasi	computer security	computer network
51	memahami model enkripsi & <i>security data</i>	computer security	computer security
52	<i>hacker</i> & keamanan	computer security	microcomputers
53	ancaman internet hacking dan trik menanganinya	computer security	computer security
54	uml distilled, edisi 3: panduan singkat bahasa pemodelan objek standar	computer science	computer science
55	aneka tool pembersih hard disk	computer science	computer security
56	mikroprosesor dan <i>interfacing</i>	computer science	operating systems
57	facebook	internet	internet
58	search engine <i>optimization</i> : cara cepat mendapatkan rating tinggi di search engine	internet	internet
59	google cheat	internet	internet

No	Judul	Klasifikasi oleh Perpustakaan Undika	Kategori sistem
60	teknik jitu memfilter internet	<i>internet</i>	<i>internet</i>
61	3 2 1 proteksi gratis saat anak ber-internet	<i>internet</i>	<i>internet</i>
62	<i>coreldraw</i> x4	<i>computer graphics</i>	<i>computer graphics</i>
63	kreasi desain flora dengan <i>coreldraw</i> x3	<i>computer graphics</i>	<i>computer graphics</i>
64	desain pop art untuk orang awam	<i>computer graphics</i>	<i>computer graphics</i>
65	115 teknik profesional <i>corel draw</i> x4	<i>computer graphics</i>	<i>computer graphics</i>
66	110 trik rahasia corel x4 : mengungkap dan menggali rahasia tersembunyi <i>software</i> grafik populer corel x4	<i>computer graphics</i>	<i>computer graphics</i>
67	data nomor pokok perpustakaan (npp) perpustakaan khusus dan sekolah se jawa timur tahun 2015	<i>library</i>	<i>library</i>
68	matabaru penelitian perpustakaan dari serqual ke libqual	<i>library</i>	<i>library</i>
69	ragam inspirasi perpustakaan rumah	<i>library</i>	<i>library</i>
70	dasar - dasar operations research, edisi 2	<i>operations research</i>	<i>operations research</i>
71	teori dan soal-soal operations research (<i>schaum s series</i>)	<i>operations research</i>	<i>operations research</i>
72	<i>microprocessor databook</i>	<i>microprocessors</i>	<i>microprocessors</i>
73	dari <i>chip</i> ke sistem : pengantar mikroprosesor	<i>microprocessors</i>	<i>microprocessors</i>
74	<i>artificial intelligence</i>	<i>artificial intelligence</i>	<i>data communication</i>
75	mengenai <i>artificial intelligence</i>	<i>Artificial Intelligence</i>	<i>computer science</i>
76	analisa dan perancangan sistem pengolahan data	<i>system analysis</i>	<i>system analysis</i>
77	komunikasi data	<i>data communication</i>	<i>data communication</i>
78	pengolah mikro/komputer mikro (seri pelajaran elektronika)	<i>MicroComputers</i>	<i>computers</i>

4.4 Evaluasi

Berikut hasil evaluasi yang didapatkan dari implementasi perancangan sistem klasifikasi kategori. Untuk menentukan Akurasi klasifikasi akan menggunakan rumus berikut :

$$Akurasi = \frac{\text{Jumlah Kategori Benar}}{\text{Jumlah dokumen uji}} \times 100\%$$

$$Akurasi = \frac{62}{78} \times 100\%$$

$$Akurasi = 79.48\% = 79\%$$

Berdasarkan Pengujian Akurasi diatas maka, didapatkan hasil kategori yang sesuai adalah sejumlah 62 dan kategori yang tidak sesuai sejumlah 16 dan hasil akurasi klasifikasi kategori menggunakan algoritma NBC sebesar 79%. Hal ini menunjukkan bahwa algoritma NBC dapat digunakan sebagai algoritma untuk pengklasifikasian kategori buku karena tingkat akurasinya yang cukup tinggi. Persentase uji akurasi akan semakin tinggi jika penggunaan data *Training* semakin banyak dan uji akurasi yang didapatkan diatas berdasarkan data *Training* sejumlah 311 data dengan lama waktu untuk menyiapkannya yaitu 10 menit 11 detik. Dikarenakan pada penelitian ini menggunakan perbandingan data kamus kategori maka banyaknya data pada kamus kategori juga dapat mempengaruhi nilai akurasi dari hasil pengujian.

Lama waktu klasifikasi kategori dengan sistem hanya membutuhkan 2 detik 70 milidetik yang mana sebelumnya ketika dilakukan secara manual klasifikasi kategori dapat memakan waktu 3 sampai 10 menit. Hal ini menunjukkan bahwa klasifikasi kategori dengan sistem klasifikasi kategori menggunakan algoritma NBC dapat dilakukan menjadi lebih cepat dan efektif.

Sistem yang sudah dirancang telah didemokan kepada pihak Perpustakaan Undika dan menghasilkan beberapa saran perbaikan dan pengembangan. Perbaikan yang sudah dilakukan yaitu menampilkan nomor DDC dan penggunaan font yang informative. Pengembangan sistem disarankan untuk membuat inputan kategori baru agar lebih fleksibel digunakan. Sesuai hasil dari google form pada lampiran 5 hasil *Google Form*. Tingkat kepuasan dari penerimaan sistem menunjukkan sesuai/baik yaitu 43%, cukup sesuai/baik yaitu 43%, dan belum sesuai/baik yaitu 14%. Dari persentase menunjukkan hasilnya yaitu 43% sesuai atau cukup sesuai.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, sistem ini dapat melakukan pengklasifikasian buku dengan memasukan judul atau deskripsi buku berdasarkan DDC 23 pada kelas 000 yaitu *Computer science, information & general works* diproses menggunakan tahapan *text preprocessing* dengan perbandingan kamus kategori dan algoritma NBC.

1. Hasil pengujian akurasi yaitu sebesar 79% yang menunjukkan bahwa kesesuaian hasil klasifikasinya cukup tinggi.
2. Klasifikasi kategori dilakukan dalam waktu yang lebih cepat dengan sistem hanya membutuhkan 2 detik 70 milidetik.

5.2 Saran

Pada pengembangan selanjutnya terhadap sistem klasifikasi kategori pada Perpustakaan Undika terdapat beberapa saran yaitu :

1. Pengembangan Sistem klasifikasi kategori dapat dilanjutkan menggunakan kategori sesuai DDC23 pada kelas – kelas berikutnya yaitu kelas 100 - 900.
2. Membuat daftar kamus kategori dengan jumlah lebih banyak.
3. Mengumpulkan data *Training* dari data perpustakaan dengan jumlah yang seimbang pada masing masing kategori yang dipilih dan menambah jumlah data *Training* untuk meningkatkan akurasi hasil pengujian sistem.
4. Pengembangan sistem agar lebih fleksibel untuk menambahkan kategori baru.

DAFTAR PUSTAKA

- Ariata. (2019, may 26). *Apa Itu GitHub*. Retrieved September 5, 2020, from Hostinger: <https://www.hostinger.co.id/tutorial/apa-itu-github/>
- Darmono. (2001). Pengorganisasian Koleksi Perpustakaan. In Darmono, *Perpustakaan Sekolah* (p. 114). Jakarta: Grasindo.
- Ginting, S. L., & Trinanda, R. P. (2013). Penggunaan Metode NBC Pada Aplikasi Perpustakaan. *Jurna Teknik Komputer Unikom*, 1-6.
- Han, J., Kamber, M., & Pei, J. (2012). Introduction. In J. Han, M. Kamber, & J. Pei, *Data Mining Concepts and Techniques* (pp. 6,33-34). USA: Morgan Kaufmann.
- Indriyono, B. V., Utami, E., & Sunyoto, A. (2015). Klasifikasi Jenis Buku Berdasarkan Judul dan Sinopsis Menggunakan Metode NBC (Studi kasus : STMIK Kadiri). 476-485.
- Kusuma, A. p., & Srirahayu, I. (2016). Sistem Pencarian Katalog Buku Menggunakan Metode NBC (NBC) Pada Aplikasi Mulia-Bookstore Berbasis Android. *Jurnal Antivirus*, 36-49.
- Librian, A. (2017, February 21). *Sastrawi*. Retrieved September 6, 2020, from Github: <https://github.com/sastrawi/sastrawi/blob/master/README.en.md>
- Maricar, M. A., & Pramana, D. (2019). Perbandingan Akurasi Naïve Bayes dan K-Nearest Neighbor pada Klasifikasi untuk Meramalkan Status Pekerjaan Alumni ITB STIKOM Bali. *Jurnal Sistem dan Informatika (JSI)*, 16-21.
- Nurhasan, F., Hikmah, N., & Utami, D. Y. (2018). Perbandingan Algoritma C4.5, KNN, dan Naïve Bayes untuk Penentuan Model Klasifikasi Penanggung Jawab BSI Entrepreneur Center. *Jurnal PILAR Nusa Mandiri*, 169-173.
- Pressman, R. S., & Maxim, B. R. (2015). Software Engineering: A Practitioner's Approach. In R. S. Pressman, & B. R. Maxim, *Software Engineering: A Practitioner's Approach* (pp. 41-43). New York: McGraw-Hill Education.
- Rizqiyani, v., Mulwida, A., & Putri, R. D. (2017). Klasifikasi Judul Buku dengan Algoritma Naive Bayes dan Pencarian Buku pada Perpustakaan Jurusan Teknik Elektro. 60-65.
- Setiawan, A., Astuti, I. F., & Kridalaksana, A. H. (2015). Klasifikasi dan Pencarian Buku Referensi Akademik Menggunakan Metode Naive Bayes Classifier (NBC) (Studi Kasus: Perpustakaan Daerah Provinsi Kalimantan Timur). *Jurnal Informatika Mulawarman*, 1-10.