

1. Judul Proposal

“Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C.45 (Studi Kasus : STMIK DUTA BANGSA (atau yang lain Boleh mas)) “

2. Latar Belakang

Pertumbuhan yang sangat pesat dari akumulasi data telah menciptakan kondisi kaya akan data tapi minim informasi. *Data Mining* merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data dalam jumlah besar yang diharapkan dapat mengatasi kondisi tersebut. *Data Mining* sendiri memiliki beberapa teknik salah satunya klasifikasi. Teknik klasifikasi terdiri beberapa metode, dan *decision tree* adalah bagian dari *metode* klasifikasi. Kemudian *metode decision tree* memiliki algoritma, algoritma C4.5 adalah salah satu dari algoritma yang memiliki *decision tree*.

Program Sarjana (S1) jurusan teknik komputer STMIK Duta Bangsa adalah program pendidikan akademik setelah pendidikan menengah, yang memiliki beban sekurang – kurangnya 144(sesuaikan pada kampus) (seratus empat puluh empat) sks (satuan kredit semester) yang dijadwalkan untuk 8 (delapan) semester dan dapat ditempuh dalam waktu kurang dari 8 (delapan) semester paling lama 14 (empat belas) semester. Hal ini menunjukkan bahwa masih banyak mahasiswa Program Sarjana (S1) regular di jurusan Teknik Komputer yang menempuh lama studi dari 8 semester dari yang dijadwalkan 8 semester.

Jika masa studi mahasiswa dapat diketahui lebih dini maka pihak jurusan teknik komputer dapat melakukan tindakan-tindakan yang diperlukan supaya mahasiswa dapat lulus tepat 8 semester.

Pada penelitian ini dibuat suatu Aplikasi data mining dengan algoritma C4.5 guna untuk menganalisis kemungkinan mahasiswa lulus lebih dari 8 semester dengan melakukan klasifikasi dari kumpulan data mahasiswa yang telah lulus. Dalam merancang bangun Aplikasi data mining dengan algoritma C4.5 nantinya penulis menggunakan *Delphi Xe2* dengan database MySQL.

3. Perumusan Masalah

Berdasarkan dari latar belakang di atas dapat dirumuskan : “Bagaimana merancang dan membangun sebuah Aplikasi untuk menganalisis kemungkinan mahasiswa lulus lebih dari 8 semester dengan teknik data mining menggunakan algoritma c4.5 sekaligus *output* dan antar muka sistem yang mudah di mengerti oleh pengguna (*user*)?.”

4. Pembatasan Masalah

Agar pembahasan masalah ini dapat lebih terarah, maka perlu adanya suatu batasan masalah. Maka perlu ditetapkan batasan – batasan permasalahan yang akan dibahas, yaitu :

1. Prediksi masa studi mahasiswa berdasarkan data nilai akademik. Data – data nilai akademik yang digunakan ada 6 mata kuliah yaitu algoritma pemrograman 1, Fisika Dasar 1, Fisika dasar 2, kalkulus 1, kalkulus 2 , dan (PSK) pengantar Sistem Komputer (Terserah sampean mas mata pelajarane)

2. Sistem yang dibuat hanya bisa digunakan untuk jenjang S-1.
3. Data diproses dari angkatan 2009-2013.
4. Aplikasi ini hanya memproses klasifikasi satu siswa atau seluruh mahasiswa jurusan teknik informatika.(lainya juga boleh)
5. Aplikasi ini menggunakan *Delphi XE2*
6. Database yang digunakan adalah *My SQL*

5. Tujuan Penelitian

Tugas akhir ini bertujuan untuk mengimplementasikan data mining menggunakan algoritma c4.5 dengan teknik pohon keputusan dalam sebuah aplikasi untuk memprediksi masa studi mahasiswa.

6. Manfaat Penelitian

Manfaat penelitian dibedakan menjadi dua jenis yaitu manfaat secara teoritis dan manfaat secara praktis

6.1 Manfaat Teoritis

Penelitian ini dapat digunakan sebagai referensi bagi pengembangan penelitian sejenis dimasa yang akan datang.

6.2 Manfaat secara Praktis

a Manfaat bagi akademik

Bagi akademik, manfaat yang dapat diperoleh dari proses pembuatan proposal skripsi ini adalah berupa sumbangsih laporan buku proposal skripsi, yang di harapkan mampu digunakan dengan sebaiknya, dan sebagai tambahan buku di perpustakaan

b Bagi Mahasiswa

Bagi mahasiswa, makalah ini dapat dijadikan sebagai sumber referensi maupun pedoman bagi mahasiswa lain yang mengambil penelitian sejenis.

c Bagi Universitas

Mampu memprediksi masa studi mahasiswa pada program sarjana (S1) dengan jurusan Teknik Informatika(atau lainnya).

7. Tinjauan Pustaka

Angga Ginanjar Mabror (2011) dalam Skripsi yang berjudul “ Penerapan Data Mining Di Bidang Marketing Untuk Memprediksi Potensi Kriteria Nasabah Menggunakan Metode Decision Tree Di PD BPR Kabupaten Bandung Cabang Batujajar “ *metode* yang digunakan dalam penulisan ini adalah Decision Tree dan algoritma yang dipakai adalah algoritma C45, Adapun data yang diproses dalam tugas akhir ini data angsuran nasabah kredit PD BPR Kabupaten Bandung Cabang Batujajar pada bulan juni 2009. Hasil dari proses mining aplikasi adalah terbentuknya pola data (Data patern) berupa rules atau aturan prediksi mengenai kriteria nasabah berpotensi dan tidak berpotensi untuk melakukan peminjaman kredit. Perancangan system menggunakan Delphi 7 dengan databasenya adalah MySQL.

Astrid Darmawan (2012) dalam skripsi yang berjudul “ Pembuatan Aplikasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma K- Naerest Neighborhood”. Data Mining merupakan serangkaian

proses untuk menggali nilai tambah berupa informasi yang selama ini tidak diketahui secara manual dari suatu database data mining juga memanfaatkan pengalaman atau bahkan kesalahan dimasa lalu untuk meningkatkan kualitas dari model maupun hasil analisisnya, salah satunya dengan kemampuan pembelajaran yang dimiliki teknik data mining yaitu klasifikasi. Klasifikasi merupakan tugas pembelajaran yang memetakan sebuah objek baru kedalam salah satu label class atau kategori pada objek lama yang telah didefinisikan sebelumnya. Klasifikasi ini menggunakan salah satu metode algoritma data mining yaitu K-Nearest Neighborhood (k-NN). Algoritma K-NN bekerja berdasarkan jarak kedekatan antara dua objek dengan menentukan nilai K. Nilai K merupakan parameter untuk menentukan jarak dekat antara objek baru dengan objek lama. Dengan menggunakan teknik data mining tersebut maka di perguruan tinggi dapat memanfaatkan data akademik mahasiswa yaitu Indeks Prestasi (IP) untuk memprediksi masa studi mahasiswa. Dalam aplikasi data mining ini terdiri dari kata testing dan training dengan inputan NIM dan nilai K yang terbaik, tergantung jumlah data yang digunakan. Ukuran nilai K yang besar untuk memprediksi masa studi mahasiswa belum tentu menjadi nilai K terbaik dengan tingkat keberhasilan yang tinggi begitupun sebaliknya. Sehingga diharapkan hasil akhir dari aplikasi data mining ini dapat menghasilkan masa studi mahasiswa.

Ni G.A.P. Harry Saptarini (2012) dalam Jurnal Matrix Vol.2,No.2 melakukan penelitian dengan judul “Penggunaan Algoritma C4.5 dan Logika *Fuzzy* untuk klasifikasi talenta karyawan”. dalam penelitian ini menjelaskan dalam Analisa kemampuan SDM yang dilakukan dengan mengklasifikasikan

talenta yang dimiliki oleh seorang karyawan. Adapun metode klasifikasi yang digunakan adalah Algoritma C4.5, dimana Algoritma C4.5 konvensional menggunakan data input berupa *crisp*. Dalam penelitian ini digunakan data input berupa *linguistic term* karena penilaian talenta diekspresikan menggunakan bahasa (*linguistic term*) dan dalam bentuk himpunan tidak tegas (*fuzzy*). Untuk menghasilkan data input dalam bentuk *fuzzy* dilakukan proses fuzzifikasi pada *processing* data selanjutnya data hasil *processing* akan digunakan untuk membangun pohon keputusan menggunakan Algoritma C4.5 yang kemudian proses ini dinamakan *fuzzy C4.5*.

Perbedaan penelitian yang penulis lakukan dengan penelitian sebelumnya adalah penelitian dengan judul “Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C.45” . Algoritma C4.5 dapat diimplementasikan untuk memprediksi masa studi mahasiswa dan pohon keputusan mengklasifikasikan data mahasiswa terdahulu dengan variable data nilai matakuliah pemrograman 1, Fisika Dasar 1, Fisika dasar 2, kalkulus 1, kalkulus 2 , dan (PSK) pengantar Sistem Komputer (Terserah sampean mas mata pelajarane). Informasi yang dihasilkan berupa hasil prediksi masa studi mahasiswa dengan skema pohon keputusan dibuat dengan algoritma C4.5.

8. Landasan Teori

1. Data Mining

Terdapat beberapa teknik data mining yang sering disebut-sebut dalam literatur. Namun ada 3 teknik data mining yang populer (Santosa 1999), yaitu :

1) *Association Rule Mining*

Association Rule Mining adalah teknik mining untuk menemukan asosiatif antara kombinasi atribut. Contoh dari aturan asosiatif dari analisa pembelian di suatu pasar swalayan dapat mengatur penempatan barangnya atau merancang strategi pemasaran dengan memakai kupon diskon untuk kombinasi barang tertentu.

2) *Clustering*

Berbeda dengan association rule mining dan klasifikasi dimana kelas data telah ditentukan sebelumnya, clustering dapat dipakai untuk memberikan label pada kelas data yang belum diketahui. Karena itu clustering sering digolongkan sebagai metode *unsupervised learning*. Prinsip *clustering* adalah memaksimalkan kesamaan antar cluster. *Clustering* dapat dilakukan pada data yang memiliki beberapa atribut yang dipetakan sebagai ruang multidimensi.

3) Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, pendapatan rendah.

2. Pohon Keputusan

Pohon keputusan adalah salah satu metode klasifikasi yang paling populer karena mudah di interpretasi manusia. Pohon keputusan adalah model prediksi menggunakan struktur pohon atau struktur berhirarki. Konsep dari pohon

keputusan adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan.

Data dalam pohon keputusan biasanya dinyatakan dalam bentuk tabel dengan atribut dan *record*. Atribut menyatakan suatu parameter yang dibuat sebagai kriteria dalam pembentukan *tree*. Misalkan untuk menentukan main tenis, kriteria yang digunakan adalah cuaca, angin, iklim dan temperatur. Manfaat utama menggunakan pohon keputusan adalah kemampuannya untuk membreak *down proses* pengambilan keputusan yang kompleks menjadi lebih simpel sehingga pengambilan keputusan akan menjadi lebih menginterpretasikan solusi permasalahan. Pohon keputusan juga berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan sebuah variabel target. Pohon keputusan memadukan antara eksplorasi data dan pemodelan sehingga sangat bagus sebagai langkah awal pemodelan bahkan ketika dijadikan sebagai model akhir dari beberapa teknik lain.

3. Algoritma C4.5

Banyak algoritma yang dapat dipakai dalam pembentukan pohon keputusan, antara lain ID3, CART, dan C4.5. algoritma C4.5 merupakan pengembangan dari algoritma ID3.

Penulis memilih algoritma C4.5 dalam pembuatan aplikasi data mining penerimaan calon pegawai karena algoritma C4.5 merupakan algoritma yang digunakan untuk melakukan pembentukan pohon keputusan. Pohon keputusan tersebut mampu menghasilkan keputusan kompleks menjadi lebih sederhana,

sehingga pengambil keputusan akan lebih menginterpretasikan solusi dari permasalahan.

Secara umum algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut.

- 1) Pilih atribut sebagai akar.
- 2) Buat cabang untuk tiap-tiap nilai.
- 3) Bagi kasus dalam cabang.
- 4) Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama.

Untuk memilih atribut sebagai akar, didasarkan pada nilai gain tertinggi dari atribut-atribut yang ada. Untuk menghitung gain digunakan rumus seperti tertera sebagai berikut :

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n * Entropy(S_i)$$

Keterangan :

S : himpunan kasus

A: Atribut

n : jumlah partisi atribut

|S_i| : jumlah kasus pada partisi ke –i

|S| : jumlah kasus dalam S

Setelah mendapatkan nilai gain, ada satu hal lagi yang perlu kita lakukan perhitungan yaitu mencari nilai Entropy. Entropy digunakan untuk menentukan seberapa informative sebuah input atribut untuk menghasilkan output atribut.

Rumus dasar dari entropy tersebut adalah sebagai berikut :

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i$$

Keterangan :

S : Himpunan kasus

A : Fitur

n : Jumlah partisi S

p_i : Proporsi dari S_i terhadap S

4. DATABASE

Sebuah database adalah sebuah struktur yang umumnya dikategorikan dalam 2 hal: Sebuah database flat dan sebuah database relasional. Database relasional lebih disukai karena lebih masuk akal dibandingkan *database flat* *MySQL* adalah sebuah *database* relasional.

Pada *database* yang memiliki struktur relasional. Ada tabel-tabel yang menyimpan data. Setiap tabel terdiri dari kolom dan baris. Sebuah kolom mendefinisikan jenis informasi apa yang akan disimpan. Diperlukan kolom khusus untuk setiap jenis informasi yang ingin di simpan (misalnya umur, tinggi, berat, alamat).(Abdul Kadir:2002).

Kalau kolom mendefinisikan jenis informasi apa yang akan disimpan, maka sebuah baris adalah data aktual yang disimpan. Setiap baris dari tabel adalah masukan dari tabel tersebut dan berisi nilai-nilai untuk setiap kolom tabel tersebut. (contoh masukannya adalah Jojon: umur 35 tahun, 65 kg dan 165cm).

5. PEMROGRAMAN BORLAND DELPHI

4.1. Pengertian Delphi

Delphi merupakan Suatu bahasa pemrograman (development language) yang digunakan untuk merancang suatu aplikasi program. (Madcoms, 2002).

4.2. Kegunaan Delphi

1. Untuk membuat aplikasi *windows*
2. Untuk merancang suatu aplikasi program berbasis grafis
3. Untuk membuat program berbasis jaringan (*client/server*)
4. Untuk merancang program .NET (berbasis *internet*)

4.3. Keunggulan Delphi

1. IDE (*Intregated Development Environment*) atau lingkungan perkembangan aplikasi sendiri adalah satu dari keunggulan delphi, didalamnya terdapat menu-menu yang memudahkan kita untuk membuat suatu proyek program

2. Proses kompilasi cepat, pada saat aplikasi yang kita buat dijalankan pada Delphi, maka secara otomatis akan dibaca sebagai sebuah program, tanpa dijalankan terpisah
3. Mudah digunakan, *source* kode delphi yang merupakan turunan dari *Pascal*, sehingga tidak diperlukan suatu penyesuaian lagi.
4. Bersifat *multi purpose*, artinya bahasa pemrograman *Delphi* dapat digunakan untuk mengembangkan berbagai keperluan pengembangan aplikasi.

4.4. Sejarah Borland Delphi

5. *Delphi* versi 1 (Berjalan pada windows 3.1 atau windows 16 bit)
6. *Delphi* versi 2 (Berjalan pada windows 95 atau delphi 32 bit)
7. *Delphi* versi 3 (Berjalan pada windows 95 keatas dengan tambahan fitur *internet* atau web)
8. Perkembangan selanjutnya diikuti dengan *Delphi* versi 4,5,6
9. Versi terkini dari *Delphi* adalah versi 7 dengan tambahan fitur .net dengan tambahan file XML

6. MYSQL

MySQL (*My Structure Query Language*) atau yang biasa di baca “mai-es-kuel” adalah sebuah program pembuatan database yang bersifat *open source*, artinya siapa saja boleh menggunakannya dan tidak akan di cekal (Bunafit Nugroho: 2004).

MySQL AB membuat MySQL tersedia sebagai perangkat lunak gratis dibawah lisensi GNU *General Public License* (GPL), tetapi mereka juga menjual dibawah lisensi komersial untuk kasus-kasus dimana penggunaannya tidak cocok dengan penggunaan GPL. MySQL adalah *Relational Database Management System* (RDBMS) yang didistribusikan secara gratis dibawah lisensi GPL (*General Public License*). Dimana setiap orang bebas untuk menggunakan MySQL, namun tidak boleh dijadikan produk turunan yang bersifat *closed source* atau komersial. MySQL sebenarnya merupakan turunan salah satu konsep utama dalam database sejak lama, yaitu SQL (*Structured Query Language*). SQL adalah sebuah konsep pengoperasian database, terutama untuk pemilihan atau seleksi dan pemasukan data, yang memungkinkan pengoperasian data dikerjakan dengan mudah secara otomatis.

MySQL adalah sebuah *server database open source* yang termasuk populer keberadaannya. MySQL umumnya digunakan bersamaan dengan PHP untuk membuat aplikasi *server* yang dinamis dan powerfull.

9. Metodologi Penelitian

Dalam melakukan penelitian ini penulis menggunakan beberapa metode, dalam membuat penelitian ini antara lain :

a. Jenis Dan Sumber Data

Selama penelitian di **STMIK DUTA BANGSA (atau lainnya)**. Penulis dapat mengumpulkan beberapa data, antara lain:

- 1) Data primer adalah data yang secara langsung diambil dari objek-objek penelitian oleh peneliti perorangan maupun organisasi. Penulis disini memperoleh data dengan mewawancarai langsung dari beberapa petugas di bagian gudang, bagian penjualan, bagian keuangan. Data sekunder adalah data yang didapat secara tidak langsung dari objek penelitian. Penulis disini mendapatkan data dari studi literatur.

b. Metode pengumpulan data

Metode pengumpulan data yang penulis gunakan adalah sebagai berikut:

- 1) Metode *Interview* atau Wawancara

Metode *interview* atau wawancara yaitu mencari secara langsung dari pihak yang bersangkutan atau *interview* langsung. Dengan mengadakan tanya jawab secara langsung kepada **Bapak Abdulah Azzam, S.Kom.** sebagai direksi yang ada kaitanya dengan topik yang diambil.

- 2) Metode Observasi

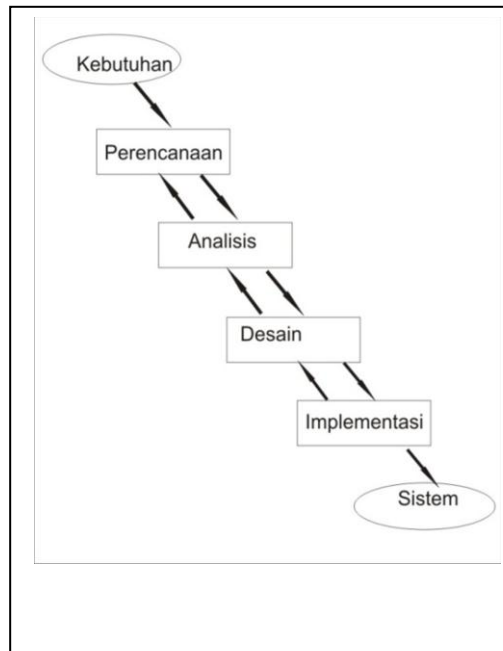
Metode observasi merupakan suatu metode pengambilan data dengan cara mengamati langsung tempat atau objek yang akan diteliti. Dalam penelitian ini yang menjadi objek penelitian yaitu **di CV. Dinamika Ilmu Surakarta**

- 3) Studi literatur

Studi ini dimaksudkan untuk pengumpulan dan memperoleh data sekunder dengan cara mempelajari, membaca dan mencatat literatur dari beberapa buku yang berkaitan dengan permasalahan di atas.

c. Metode Pengembangan Sistem

1. Prototipe adalah suatu pendekatan yang sangat rapi dan berurutan untuk membuat sebuah sistem menjadi sesuatu yang nyata.
2. Prototipe mempunyai proses design berulang dan mengkombinasikan empat fase utama SDLC(System Development Life Cycle) tradisional (analisi,desain,kontruksi dan implementasi) ke dalam satu langkah yang berulang-ulang.



Gambar 2 *SDLC* tradisional
(Sumber: Efraim Turban,dkk :2005:402)

1) Analisis

Langkah ini merupakan analisa terhadap kebutuhan sistem. Pengumpulan data dalam tahap ini dapat melakukan sebuah penelitian, wawancara atau *study literatur*. Seorang sistem analis akan menggali informasi sebanyak-

banyaknya dari *user* sehingga akan tercipta sebuah sistem komputer yang mampu melakukan tugas-tugas yang diinginkan oleh *user* tersebut. Tahapan ini akan menghasilkan dokumen *user requirment* atau dapat dikatakan sebagai data yang berhubungan dengan keinginan *user* dalam pembuatan sistem. Dokumen ini lah yang akan menjadi acuan sistem analis untuk menterjemahkan ke dalam bahasa pemrogram. Fase ini harus dikerjakan secara lengkap untuk dapat menghasilkan *design* yang lengkap.

- a) Analisis kelemahan sistem menggunakan metode *Pieces*.
- b) Analisis kebutuhan sistem: analisis kebutuhan informasi, analisis kelayakan sistem.

2) Design

Desain adalah proses multi langkah yang fokus pada desain pembuatan program perangkat lunak.

- a) Desain antarmuka: desain struktur menu, desain login, desain menu utama, desain input, dan desain laporan.
- b) Alat bantu perancangan menggunakan *Flowchart*

3) Coding dan Testing

Coding merupakan penerjemahan *design* dalam bahasa yang dapat dikenali oleh komputer. Dilakukan oleh *programmer* yang akan meterjemahkan transaksi yang diminta oleh *user*. Tahapan ini yang merupakan tahapan secara nyata dalam mengerjakan suatu sistem. Penulis dalam pengkodean menggunakan bahasa pemograman *Delphi* dan *database MySQL*. Setelah

pengkodean selesai maka akan dilakukan *testing* terhadap sistem yang telah dibuat tadi. Tujuan *testing* adalah menemukan kesalahan-kesalahan terhadap sistem tersebut dan kemudian dapat diperbaiki.

4) Penerapan

Tahapan ini dapat dikatakan *final* dalam pembuatan sebuah sistem. Setelah melakukan analisa, *design* dan pengkodean maka sistem yang sudah jadi akan digunakan oleh *user*.

5) Pemeliharaan

Perangkat lunak yang sudah disampaikan kepada pelanggan pasti akan mengalami perubahan. Perubahan tersebut mampu karena mengalami kesalahan karena perangkat lunak harus menyesuaikan dengan lingkungan (*peripheral* atau sistem operasi baru) baru, atau karena pelanggan membutuhkan perkembangan fungsional.

10. Lokasi Penelitian

**PT. CV. DINAMIKA ILMU Jl. Klengkeng No. 03 Kerten Rt.02/08
Surakarta (0271) 731563**

11. Sistematika Penulisan

Sistematika yang akan diajukan adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang, perumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, tinjauan pustaka, metodologi penelitian dan sistematika

penulisan.

BAB II LANDASAN TEORI

Dalam bab ini berisi tentang landasan teori yang digunakan dalam mendukung penyusunan laporan skripsi. Seluruh teori-teori yang mendukung dalam laporan skripsi disajikan dalam landasan teori.

BAB III ANALISIS DAN PERANCANGAN SISTEM

Dalam bab ini berisi tentang identifikasi masalah, analisis, solusi alternatif dan desain rancangan sistem secara umum.

BAB IV IMPLEMENTASI SISTEM

Dalam bab ini disajikan pembahasan tentang bagaimana sistem diterapkan di **STMIK DUTA BANGSA (ATAU LAINYA)**

BAB V PENUTUP

Bab ini berisi tentang rangkuman mengenai hasil pokok yang telah dibahas pada bab-bab sebelumnya atau kesimpulan dan saran.

12. Jadwal Pelaksanaan

Jadwal penelitian dibuat sesuai dengan waktu yang dibutuhkan untuk penelitian dan disesuaikan dengan metodologi penelitian yang digunakan.

No	Kegiatan	Bulan															
		Mei				Juni				Juli				Agustus			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4

[illegible]