

Implementasi Algoritma C4.5 Dalam Memprediksi Penjurusan Untuk Siswa SMK Negeri 1 Nanga Pinoh

Adi Triyanto¹, Masri²

Jurusan Teknik Informatika, STMIK Pontianak

STMIK Pontianak; Jl. Merdeka No.372 Pontianak, 0561-735555

e-mail: [1aditriyanto920@gmail.com](mailto:aditriyanto920@gmail.com), [2masri@stmikpontianak.ac.id](mailto:masri@stmikpontianak.ac.id)

Abstrak

Penentuan jurusan di SMK merupakan hal yang sangat penting, karena siswa akan dibekali pembelajaran sesuai dengan jurusan yang telah dipilih sehingga memiliki keterampilan sebagai bekal memasuki dunia kerja. Namun dalam prosesnya masih sering mengalami masalah antara lain, terkadang membuat bingung siswa dalam menentukan jurusan pada saat masuk SMK, minat dan bakat siswa terkadang belum terlihat sehingga siswa juga kesulitan dalam menentukan jurusan. Penelitian ini menggunakan metode penelitian Action Research, yang terdiri dari : melakukan diagnose (Diagnosing), membuat rencana Tindakan (Action Planning), melakukan tindakan (Action Taking), melakukan evaluasi (Evaluating), pembelajaran (Learning). Data yang digunakan dalam penelitian ini adalah data siswa SMK kelas X untuk membangun decision tree dan data siswa SMK berprestasi kelas X untuk mengevaluasi kinerja dari sistem. Penelitian ini menggunakan algoritma C4.5 untuk mengolah data dalam pembentukan decision tree. Penentuan atribut yang digunakan sebagai masukan dan keluaran sistem dilakukan pada proses analisis data. Hasil dari perancangan Algoritma C4.5 ini memberikan hasil prediksi yang menentukan jurusan mana yang cocok bagi siswa SMK Negeri 1 Nanga Pinoh, berdasarkan beberapa aspek penilaian yaitu menampilkan informasi mengenai para peserta didik yang mendaftar, menampilkan informasi dalam proses seleksi menggunakan nilai hasil ujian nasional. menampilkan informasi mengenai nilai Tes Penjurusan dan minat peserta didik.

Kata kunci: C4.5, klasifikasi, pohon keputusan, jurusan

Abstract

Determination of majors in SMK is very important, because students will be provided with learning in accordance with the majors that have been chosen so that they have the skills to enter the world of work. However, in the process, they often experience problems, among others, sometimes confusing students in determining majors at the time of entering Vocational High School, students' interests and talents are sometimes not visible so students also have difficulty in determining majors. This study uses Action Research research methods, which consist of: diagnosing, making action plans, taking action, evaluating, and learning. The data used in this study is data for class X SMK students to build a decision tree and data for high achieving SMK students in class X to evaluate the performance of the system. This study uses the C4.5 algorithm to process data in the formation of a decision tree. Determination of the attributes used as input and output of the system is carried out in the data analysis process. The results of the design of the C4.5 Algorithm provide predictive results that determine which majors are suitable for students of SMK Negeri 1 Nanga Pinoh, based on several aspects of the assessment, namely displaying information about students who register, displaying information in the selection process using the results of the national exam. displays information about the majoring test scores and student interests.

Keywords: C4.5, classification, decision tree, major.

1. Pendahuluan

Data mining merupakan penambangan atau penemuan informasi baru dengan mencari pola atau aturan tertentu dari sejumlah data yang sangat besar [1]. Data mining juga disebut sebagai serangkaian proses untuk menggali nilai tambah berupa pengetahuan yang selama ini tidak diketahui secara manual dari suatu kumpulan data [2].

Algoritma C4.5 merupakan pengembangan dari algoritma ID3. Algoritma C4.5 merupakan struktur pohon di mana terdapat simpul yang mendeskripsikan atribut-atribut. Atribut-atribut tersebut

merupakan parameter kriteria dalam pembentukan pohon. Sedangkan proses dalam pohon keputusan yaitu mengubah bentuk data menjadi model pohon kemudian mengubah model pohon tersebut menjadi *rule*, dan menyederhanakan *rule* [3]. Algoritma C4.5 merupakan algoritma klasifikasi pohon keputusan yang banyak digunakan karena memiliki kelebihan utama dari algoritma yang lainnya. Kelebihan algoritma C4.5 dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima, efisien dalam menangani atribut bertipe diskret dan dapat menangani atribut bertipe diskret dan numerik. [4].

Algoritma C4.5 digunakan sebagai metode data mining untuk menyelesaikan suatu permasalahan pada penelitian ini yaitu dalam memprediksi penjurusan siswa pada SMK Negeri 1 Nanga Pinoh [5]. Beberapa masalah yang timbul pada SMK Negeri 1 Nanga Pinoh tersebut antara lain, nilai pelajaran yang didapatkan di tingkat SMP, terkadang membuat bingung siswa dalam menentukan jurusan pada saat masuk SMK, minat dan bakat siswa terkadang belum terlihat sehingga siswa juga kesulitan dalam menentukan jurusan. Atribut yang menentukan penjurusan pada SMK Negeri 1 Nanga Pinoh adalah nilai rata-rata raport, nilai Ujian Nasional, nilai Tes Penjurusan dan minat peserta didik. Untuk mengatasi permasalahan tersebut, maka dari itu dibutuhkan sistem prediksi yang akan digunakan sebagai acuan peserta didik dalam menentukan jurusan yang sesuai dengan hasil prediksi tersebut [6]. Dengan membangun suatu sistem prediksi yang dapat membantu pihak sekolah untuk meningkatkan permintaan penjurusan peserta didik kelas X (sepuluh) yaitu dengan data mining dan metode yang digunakan untuk menyelesaikan permasalahan tersebut dengan algoritma C4.5. Algoritma C4.5 yang dapat melakukan proses learning dan klasifikasi yang sederhana, cepat dan secara umum mempunyai tingkat akurasi yang tinggi. [7]. Selain itu terdapat juga beberapa tinjauan penelitian yang berkaitan yaitu penelitian membahas tentang algoritma C4.5 untuk prediksi kelayakan masuk penjurusan IPA Siswa Sekolah Menengah Atas memaparkan bahwa implementasi *data mining* algoritma C4.5 dalam memprediksi kelayakan penjurusan IPA pada SMA berhasil diimplementasikan dengan baik, dan berdasarkan hasil uji coba validasi dengan menggunakan *10-cross validation* memiliki nilai tingkat *error* sebesar 16,65%. Hasil ini menunjukkan tingkat akurasi aplikasi yang dibangun adalah 83,35 yang tergolong tinggi [8].

Selain itu terdapat juga beberapa tinjauan penelitian yang berkaitan untuk dapat mengklasifikasikan penjurusan siswa dengan menggunakan algoritma C4.5 yang akan diproses secara otomatis tanpa menggunakan perangkat lunak machine learning untuk menghasilkan penilaian tingkat kinerja hasil klasifikasi penjurusan siswa dengan menggunakan tabel penelitian *confusion matrix* menghasilkan nilai *accuracy* sebesar 89.74%, nilai *precision* sebesar 96.55% , dan nilai *recall* sebesar 90.32% [9].

2. Metode Penelitian

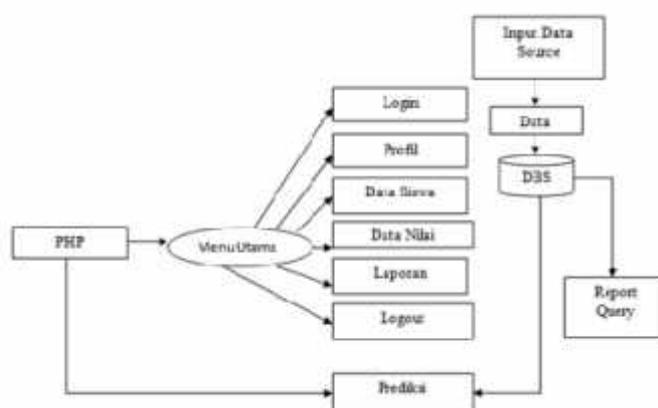
Bentuk penelitian yang digunakan oleh penulis pada penelitian ini adalah studi kasus. Penulis mempelajari secara intensif latar belakang dan kondisi pada saat ini yang dialami dalam memprediksi penjurusan siswa pada SMK Negeri 1 nanga pinoh. Studi kasus dilakukan dengan melakukan wawancara dengan Kepala Sekolah SMK Negeri 1 nanga pinoh mengenai latar belakang sekolah, masalah yang ditemukan dan data-data mengenai penjurusan siswa pada SMK Negeri 1 Nanga Pinoh. Metode pengumpulan data merupakan bagian paling penting dalam sebuah penelitian dimana ketersediaan data dapat mendukung suatu analisis maupun penelitian atas suatu obyek yang diteliti. Metode pengumpulan data yang digunakan peneliti ini adalah menggunakan metode pengumpulan data primer dan metode pengumpulan data sekunder. Data Primer ini didapat dari Kepala Sekolah SMK Negeri 1 nanga pinoh, data tersebut berupa data siswa, nilai rapor, nilai ujian ahir sekolah dan nilai ujian akhir nasional. Data Sekunder yang yang diperoleh dari Kepala Sekolah SMK Negeri 1 nanga pinoh berupa sejarah Sekolah SMK Negeri 1 Nanga Pinoh.

Teknik pengumpulan data adalah cara-cara yang dilakukan untuk mengumpulkan, mencari dan memperoleh data yang dibutuhkan dalam melakukan penelitian, baik data itu diperoleh dengan survei langsung maupun dengan penggalian informasi. Teknik pengumpulan data yang digunakan adalah sebagai berikut, teknik pengumpulan data dengan cara melakukan wawancara langsung atau aktivitas tanya jawab secara dua arah untuk mendapat data dari responden. Instrumen penelitian dipilih dan digunakan oleh peneliti dalam rangka pengumpulan data yang membantu dalam proses penelitian sesuai dengan variabel yang telah ditetapkan dalam penelitian. Penulis menggunakan alat bantu berupa perangkat elektronik berupa smartphone yang digunakan untuk merekam. Variabel yang digunakan peneliti adalah variabel tunggal yaitu perancangan perangkat lunak untuk prediksi penjurusan SMK Negeri 1 Nanga Pinoh dengan menerapkan algoritma C4.5 untuk menentukan penjurusan siswa SMK Negeri 1 Nanga Pinoh.

Metode penelitian yang di gunakan pada penelitian ini adalah metode *Action research*. *Action research* adalah sebagian sebuah metode penelitian yang didirikan atas asumsi bahwa teori dan praktik dapat secara tertutup diintergrasikan dengan pembelajaran hasil intervensi yang direncanakan setelah diagnosis yang rinci terhadap konteks masalah. Metode perancangan perangkat lunak yang di gunakan pada penelitian ini adalah metode *Extreme Programming*, adapun tahapan dari *Extreme Programming* adalah *planning, design, coding, dan testing*. Tahap pertama adalah *planning* yaitu melakukan wawancara dan observasi pada SMK Negeri 1 Nanga Pinoh setelah itu melakukan analisis terhadap kebutuhan yang diperlukan oleh SMK Negeri 1 Nanga Pinoh, tahap kedua adalah *design* yaitu melakukan *design* perangkat lunak menggunakan UML dengan empat diagram yaitu *activity diagram, class diagram, sequence diagram* dan *use case diagram*. Tahap ketiga adalah *coding* yaitu mengimplementasikan design perangkat lunak tersebut ke dalam coding menggunakan bahasa pemrograman javascript, PHP, CSS, HTML. Tahap keempat adalah *testing* yaitu melakukan testing perangkat lunak menggunakan metode *white-box*.

3. Hasil dan Pembahasan

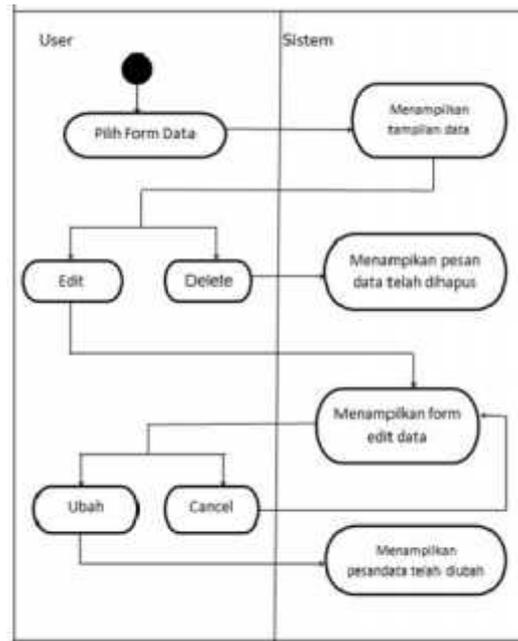
Tujuan dari perencanaan syarat-syarat dilakukan untuk melakukan pengidentifikasian tujuan-tujuan perangkat lunak atau sistem untuk mengidentifikasi syarat-syarat informasi yang ditimbulkan dari tujuan-tujuan tersebut. Peranan analisis pada tahapan ini sangatlah penting sebagai fasilitas dalam mengetahui kebutuhan pengguna. Pengumpulan data merupakan komponen yang penting dalam tahap analisis ini. Pengumpulan data dilakukan dengan mewawancarai pemilik, observasi dan studi dokumen. Penulis mendeskripsikan segala hal yang diperlukan dalam rangka pengembangan sistem. Tahap pendefinisian harus mengumpulkan kebutuhan *software* maupun hardware selengkap-lengkapnyanya. Adapun yang menjadi kebutuhan dasar dalam perancangan aplikasi desktop prediksi penjurusan siswa SMK Negeri 1 Nanga Pinoh , aplikasi desktop harus bisa menampilkan informasi penjurusan secara rinci dan mudah diakses oleh siswa. Informasi yang ditampilkan dalam aplikasi desktop harus lebih interaktif. Aplikasi desktop dirancang lebih fleksibel agar memberikan kemudahan bagi siswa untuk mendapatkan informasi dengan baik. Pastikan aplikasi desktop dapat diakses dengan baik. Perancangan pemodelan perangkat lunak yang digunakan dalam merancang sistem aplikasi desktop prediksi penjurusan siswa SMK Negeri 1 Nanga Pinoh adalah menggunakan UML (Unified Modelling Language) yang terdiri dari *use case, activity diagram, class diagram, dan squence diagram*. Berikut ini adalah rancangan pemodelan perangkat lunak aplikasi desktop prediksi penjurusan siswa SMK Negeri 1 Nanga Pinoh.



Gambar 1 Model Arsitektur Perangkat Lunak Prediksi Penjurusan Siswa SMK Negeri 1 Nanga Pinoh

3.1 Arsitektur Perangkat Prediksi Penjurusan Siswa SMK Negeri 1 Nanga Pinoh

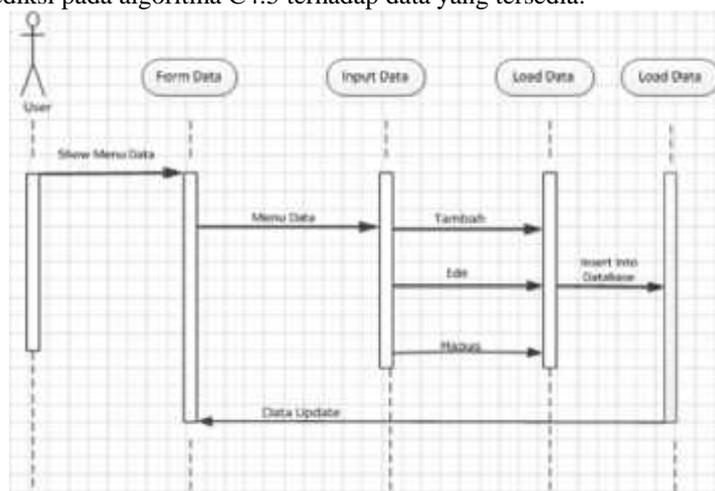
Perancangan model arsitektur mengidentifikasi semua struktur sistem hubungannya dan bagaimana didistribusikan. Pada Gambar 1, dapat dilihat bahwa aplikasi prediksi data mining sekolah dan yang menjalankan perangkat lunak di atas adalah admin dengan cara melakukan input data berdasarkan dari kedua menu yaitu data siswa dan laporan, penginputan data tersebut tersimpan didalam database. metode prediksi ini peneliti menggunakan algoritma C4.5, proses prediksi ini dilakukan pada data siswa yang diambil dari *excel* dan *database*.



Gambar 2 Activity Diagram

3.2 Activity Diagram

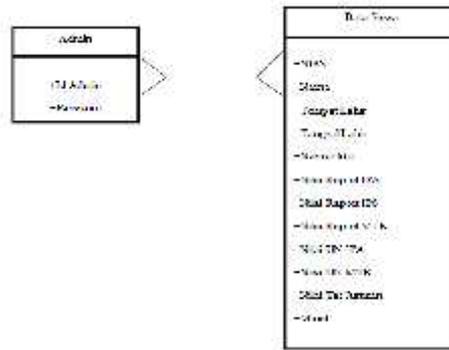
Activity Diagram adalah diagram yang menggambarkan aktivitas dari sistem yang dirancang. Activity Diagram dapat digunakan untuk menunjukkan aliran kerja dari sistem dan dapat juga digunakan untuk menggambarkan aliran dari kejadian. Node pada activity diagram disebut dengan action bukan activity. Activity menunjukan ke urutan action, sehingga diagram menunjukan activity yang membangun action. Berikut ini adalah activity diagram dari perangkat lunak aplikasi desktop prediksi penjurusan siswa SMK Negeri 1 Nanga Pinoh. Pada gambar 2 menggambarkan bahwa admin dapat melakukan prediksi dengan mnegisi data pada form poses prediksi yang ditampilkan oleh sistem terlebih dahulu. Adapun data yang harus diisi berupa prediksi yang diinginkan pengguna serta file yang berisi data mengenai objek yang akan dikelompokkan yaitu nama objek, latitude dan longitude. Setelah user mengisi data pada form proses prediksi, dan button prediksi dipilih maka sistem akan melakukan perhitungan dengan metode prediksi pada algoritma C4.5 terhadap data yang tersedia.



Gambar 3 Sequence Diagram Data

3.3 Sequence Diagram

Sequence Diagram (Gambar 3) merupakan penjelasan alur kerja sistem untuk menu data user masuk untuk melakukan pengisian data masyarakat yang akan diproses atau dicluster. Pada proses ini, jika proses penginputan datanya berhasil maka sistem akan mengupdate database sesuai dengan aksi yang dilakukan.



Gambar 4. Class Diagram

3.4 Class Diagram

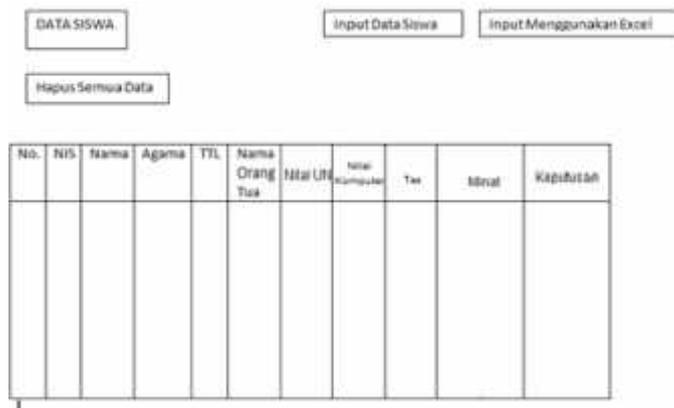
Class diagram menggambarkan struktur sistem dari segi pendefinisian kelas-kelas yang akan dibuat untuk membangun suatu sistem. berikut adalah *Class Diagram* aplikasi desktop prediksi penjurusan siswa SMK Negeri 1 Nanga Pinoh. Gambar 4 *Class diagram* ialah menghubungkan suatu *class* satu dengan *class* yang lainnya, agar suatu saat di panggil untuk melakukan inputan tidak akan muncul bersamaan, *class diagram* mempunyai tahapan untuk di panggil, sehingga inputan yang sudah dimasukkan sesuai dengan tujuan *class* masing-masing.



Gambar 5. Desain Beranda

3.5 Desain Beranda

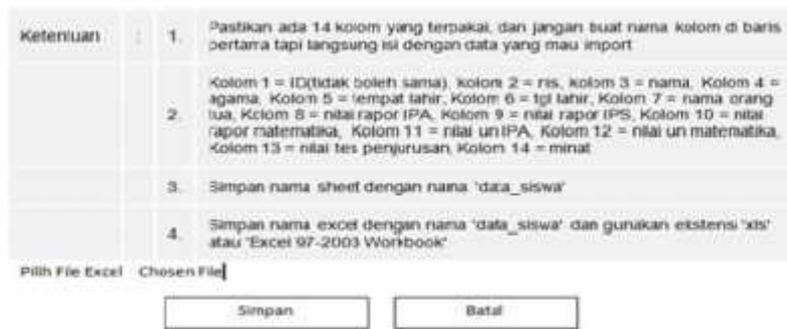
Pada Gambar 5, merupakan desain menu Beranda yang berisi kan gambar SMK Negeri 1 Nanga Pinoh beserta Visi dan Misi sekolah.



Gambar 6 Desain Data Siswa

3.6 Desain Data Siswa

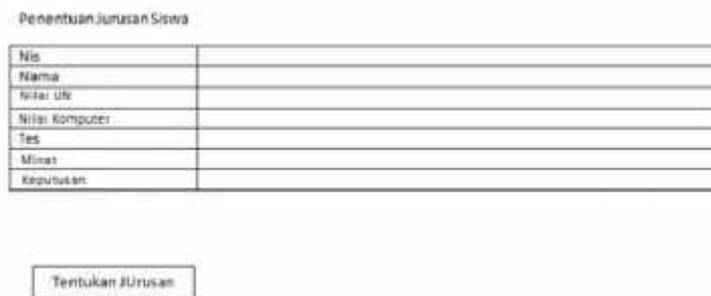
Pada Gambar 6, merupakan desain menu Data Siswa yang bisa diinputkan secara satu persatu, yang nanti akan menghasilkan sebuah keputusan dari minat yang siswa tersebut ambil.



Gambar 7 Desain Data Siswa Input Dengan Menggunakan Excel

3.7 Desain Data Siswa Input Dengan Menggunakan Excel

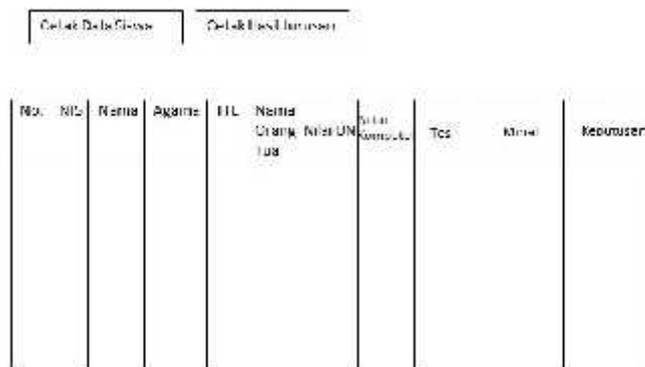
Pada Gambar 7, merupakan desain menu Data Siswa yang bisa diinputkan secara manual dengan menggunakan excel.



Gambar 8 Desain Perhitungan Algoritma C4.5

3.8 Perancangan Transaksi Penjualan

Gambar 8, merupakan desain menu Perhitungan Algoritma C4.5 yang menentukan hasil akhir siswa dari seleksi yang dilakukan.



Gambar 9 Desain Laporan

3.9 Desain Laporan

Pada Gambar 9, merupakan desain menu Desain Laporan seluruh siswa.



Gambar 10. From Data Siswa

3.10 Form Dat Siswa

Pada Gambar 10 merupakan tampilan form dari data siswa yang digunakan untuk menyimpan data siswa. Dalam form ini terdapat *button* yaitu button input data, *button* hapus data, input data menggunakan *excel*. Selain digunakan untuk penyimpanan, pada form ini dapat dilakukan pengeditan dan penghapusan data siswa yang terdapat di dalam *database*.



Gambar 11 Form Data Nilai

3.11 Form Data Nilai

Gambar 11 merupakan tampilan dari form prediksi yang digunakan untuk melihat data nilai siswa yang mencakup NISN, Nama siswa, Minat, Agama serta nilai mata pelajaran.



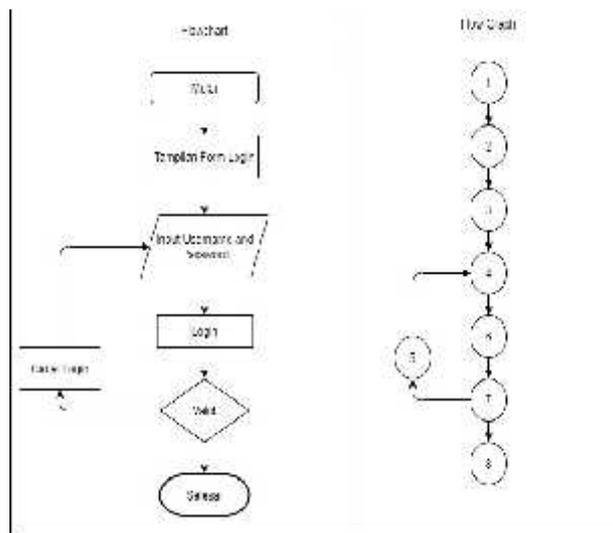
Gambar 12 Form Penentuan Jurusan

3.12 Form Penentuan Jurusan

Pada Gambar 12 merupakan tampilan dari form prediksi yang digunakan untuk memprediksi data siswa yang terdapt dalam *database*. *form* ini hanya terdapat 1 *button* yaitu tentukan jurusan yang menghasilkan perhitungan algoritma C4.5. Dalam perhitungan prediksi C4.5 ini diperlukan data yang berbentuk numerik/angka. Pada data mining pemilihan jurusan SMK Negeri 1 Nanga Pinoh ini peneliti menggunakan nilai raport SMP, nilai UN, dan nilai Tes jurusan. Namun untuk melakukan prediksi peneliti harus menentukan nilai titik awal terlebih dahulu.

3.13 Testing

Testing dan pengujian perangkat lunak adalah menguji semua elemen-elemen pada sistem perangkat lunak untuk mengetahui kesesuaian dari perancangan perancangan perangkat lunak yang dibangun. Pengujian ini dilaksanakan oleh *user* dengan metode pengujian menggunakan *White Box Testing*. Adapun pengujian *White Box Testing* gambar 13 dari rancangan perangkat lunak yang dibuat yaitu:



Gambar 13 Pengujian White-Box Testing

4. Kesimpulan

Adapun hasil perancangan perangkat lunak aplikasi desktop Prediksi Penjurusan Siswa SMK Negeri 1 Nanga Pinoh menghasilkan kesimpulan sebagai berikut:

- a. Perangkat lunak yang dirancang membantu siswa dalam memilih jurusan.
- b. Penentuan jurusan berdasarkan minat dan juga nilai siswa selama tes dilakukan.
- c. Aplikasi Desktop Prediksi Penjurusan Siswa SMK Negeri 1 Nanga Pinoh dapat memberikan kemudahan bagi siswa dalam menentukan jurusan yang diinginkan.

Daftar Pustaka

- [1] Davies. B., “*Database Systems 3rd Edition*”, Basingstoke, UK : Palgrave, 2004.
- [2] Santoso, “*Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis*”, Yogyakarta: Graha Ilmu, 2007.
- [3] Kamagi, Seng Hansun, “ Implementasi Data Mining dengan Algoritma C4.5 untuk memprediksi Tingkat Kelulusan Mahasiswa,” Universitas Multimedia Nusantara, Juni 2014.
- [4] Kamagi, Seng Hansun, “ Implementasi Data Mining dengan Algoritma C4.5 untuk memprediksi Tingkat Kelulusan Mahasiswa,” Universitas Multimedia Nusantara, Juni 2014.
- [5] Novianti, DKK. 2016. Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika. Penelitian Bidang Komputer Sains dan Pendidikan Informatika*, V2.I2 . ISSN : 2407-0491. E-ISSN : 2541-3716.
- [6] Sujai, I., Purwanto, P., & Himawan, H. (2016). Prediksi Hasil Penjurusan Siswa Sekolah Menengah Atas Dengan Menggunakan Algoritma Decision Tree C4.5. *Jurnal Cyberku*, 5-5.
- [7] Rasi, R.Z., Masroom, N.R., Omar, S.S., Ahmad, M, F., Rohana, S. (2017). *Designing Halal Supply Chain: Malaysia's Halal Industry Scenarios. MATEC Web of Conferences* 135, 00040 (2017) ICME'17, Hal 1-9.
- [8] Novianti, DKK. 2016. Data Mining : Klasifikasi Menggunakan Algoritma C4.5. *Jurnal Edik Informatika. Penelitian Bidang Komputer Sains dan Pendidikan Informatika*, V2.I2(213-219). ISSN : 2407-0491. E-ISSN : 2541-3716.