

BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi yaitu penerapan keilmuan yang mempelajari dan mengembangkan kemampuan dari suatu rekayasa dengan langkah dan teknik tertentu dalam suatu bidang. [1] Teknologi juga merupakan aplikasi ilmu dan *engineering* untuk mengembangkan mesin dan prosedur agar memperluas dan memperbaiki kondisi manusia atau paling tidak memperbaiki efisiensi manusia pada beberapa aspek. Salah satunya pada aspek dalam meningkatkan kegiatan belajar mengajar menggunakan kemajuan dari teknologi.

Ujian Nasional (UN) merupakan kegiatan pengukuran pencapaian kompetensi peserta didik pada beberapa mata pelajaran tertentu yang dilakukan di dunia pendidikan dan disesuaikan dengan standar pencapaian hasil secara nasional. Ujian Nasional atau biasa disingkat UN menjadi salah satu tolak ukur lulus tidaknya peserta didik ketika duduk di bangku sekolah dengan batas nilai minimal tertentu. Dengan adanya Ujian Nasional ini, tidak sedikit para peserta didik yang gagal untuk melewati batas nilai minimal tersebut. Hal itu disebabkan oleh tingkat kemampuan dari para peserta didik yang berbeda-beda. Perbedaan itu harus kita sadari bahwa di Negara Indonesia ini terdiri dari banyaknya suku bangsa dan budaya dengan pola kehidupan yang berbeda-beda sehingga dapat menyebabkan perbedaan pola pendidikan dan pembelajaran dari peserta didik itu sendiri.

Prediksi yaitu hasil dari kegiatan meramal atau memperkirakan suatu masalah dengan berdasarkan metode ilmiah ataupun subjektif belaka. [2] Bagus tidaknya prediksi dari suatu masalah dinilai dari sebagaimana keakuratan prediksi tersebut dengan hasil nyatanya. Selain itu terdapat banyak pula buku jenis-jenis kumpulan soal ujian nasional dalam memprediksi soal ujian yang berbeda-beda sehingga membuat para peserta ujian nasional kebingungan dalam menentukan referensi yang bagus untuk menghadapi ujian nasional. Berdasarkan masalah tersebut maka diperlukan suatu metode yang dapat membantu para peserta ujian nasional untuk meningkatkan belajarnya dalam menghadapi suatu ujian nasional. Karena pada saat ini minat untuk membaca buku semakin menurun, dan semakin beralih pada kemajuan teknologi seperti komputer, gadget dan lain sebagainya. Maka diperlukan suatu aplikasi yang dapat menunjang para peserta ujian nasional untuk mempelajari lebih dalam lagi tentang soal-soal ujian nasional yang telah dilakukan pada tahun-tahun sebelumnya. Melihat keadaan tersebut, maka dibuatkan sebuah aplikasi yang dapat memprediksi soal-soal ujian nasional pada tahun berikutnya.

Salah satu algoritma yang dapat menunjang dalam pembuatan aplikasi tersebut yaitu algoritma C 4.5. Algoritma C 4.5 merupakan pengembangan dari algoritma *Iterative Dichotomiser 3* (ID3) yaitu algoritma matematika yang digunakan untuk menghasilkan suatu pohon keputusan yang mampu mengklasifikasi suatu obyek. ID3 merepresentasi konsep-konsep dalam bentuk pohon keputusan. [3] Aturan-aturan yang dihasilkan oleh ID3 mempunyai relasi yang hirarkis seperti suatu pohon (mempunyai akar, titik, cabang, dan daun). Algoritma pada ID3 juga lebih memilih pohon keputusan yang lebih kecil (teori sederhana) dibanding yang lebih besar.

Sehingga mempunyai beberapa kekurangan yaitu jika contoh yang diteliti terlalu kecil/sederhana mungkin membuat data *over-classified*, hanya satu atribut yang dapat dites dalam satu waktu untuk membuat keputusan, dan lain sebagainya. Adapun sampel data yang digunakan oleh ID3 memiliki beberapa syarat, yaitu :

- Deskripsi atribut-nilai. Atribut yang sama harus mendeskripsikan tiap contoh dan memiliki jumlah nilai yang sudah ditentukan.
- Kelas yang sudah didefinisikan sebelumnya. Suatu atribut contoh harus sudah didefinisikan, karena mereka tidak dipelajari oleh *ID3*.
- Kelas-kelas yang diskrit. Kelas harus digambarkan dengan jelas. Kelas yang kontinyu dipecah-pecah menjadi kategori-kategori yang relative, misalnya saja mental dikategorikan menjadi "*hard, quite hard, flexible, soft, quite soft*".
- Jumlah contoh (*example*) yang cukup. Karena pembangkitan induktif digunakan, maka dibutuhkan *test case* yang cukup untuk membedakan pola yang valid dari peluang suatu kejadian.

Beberapa pengembangan yang dilakukan pada algoritma C 4.5 yaitu mampu menangani atribut dengan tipe diskrit atau kontinyu, mampu menangani atribut yang kosong (*missing value*), bisa memangkas cabang, dan pemilihan atribut yang dilakukan dengan menggunakan *Gain Ratio*. *Information Gain* pada ID3 lebih mengutamakan pengujian yang menghasilkan banyak keluaran. Dengan kata lain, atribut yang memiliki banyak nilailah yang dipilih sebagai *splitting* atribut. Karena itu algoritma C4.5 yang merupakan suksesor dari ID3 menggunakan *Gain Ratio* untuk memperbaiki *Information Gain*. Dimana atribut dengan nilai *Gain Ratio* tertinggi

dipilih sebagai atribut test untuk simpul. Dengan *Gain* sebagai *Information Gain*. Pendekatan ini menerapkan normalisasi pada *Information Gain* dengan menggunakan *Split Information* yaitu pada saat pembangunan pohon keputusan, banyaknya cabang mungkin mencerminkan adanya noise atau outlier pada training data. Pemangkasan pohon dapat dilakukan untuk mengenali dan menghapus cabang-cabang tersebut. Pohon yang dipangkas akan menjadi lebih kecil dan lebih mudah dipahami sehingga menjadi lebih cepat dan lebih baik dalam melakukan klasifikasi. Oleh karena itu Algoritma C 4.5 merupakan algoritma yang sangat tepat dan akurat untuk melakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat prediktif.

Berdasarkan pemaparan di atas, maka judul penelitian ini akan diarahkan pada **“Implementasi Algoritma C 4.5 untuk Memprediksi Soal Ujian Nasional Mata Pelajaran Matematika tingkat SMA”**. Sistem ini diharapkan mampu memberikan solusi dalam menentukan judul tugas akhir yang lebih inovatif dan berbeda dengan yang lain.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang, maka didapatkan rumusan masalah yaitu

1. Bagaimana membuat aplikasi kumpulan soal-soal ujian nasional sma beserta prediksi kemunculannya dengan menerapkan algoritma C 4.5?

1.3 Tujuan

Mengacu pada perumusan masalah di atas, tujuan dari penyusunan tugas akhir ini adalah:

- a. Membuat aplikasi kumpulan soal-soal ujian nasional sma beserta prediksinya dengan menerapkan algoritma C 4.5

1.4 Batasan Masalah

Agar pengerjaan tugas akhir ini lebih terarah, maka diperlukan pembatasan masalah. Sehingga pembahasan dibatasi pada ruang lingkup :

- a. Lingkungan Aplikasi yang dibangun berbasis *desktop*.
- b. Aplikasi yang dibangun digunakan pada koneksi *offline*.
- c. Aplikasi yang dibuat adalah aplikasi tentang soal-soal ujian nasional mata pelajaran matematika.
- d. Jawaban pada soal yang diberikan hanya jawabannya saja, tidak beserta cara penyelesaiannya.
- e. Target pengguna aplikasi ini adalah para siswa-siswa sma jurusan ipa.
- f. Algoritma yang diimplementasikan untuk memprediksi soal-soal ujian nasional sma adalah algoritma C 4.5.
- g. Metode pengembangan perangkat lunak yang diterapkan adalah *Prototype*.

1.5 The State of the Art

The State of The Art bertujuan untuk menganalisa penelitian sebelumnya yang pernah ada, yang sejalan dan mempunyai konsep yang hampir sama dengan penelitian saat ini, agar kita dapat melihat dan menyimpulkan perbedaan dari setiap penelitian sehingga masing-masing penelitian mempunyai tema dan karakter yang original. Berikut ini adalah penelitian yang telah dilakukan dan memiliki korelasi yang searah dengan penelitian yang dibahas, antara lain:

Abidin, metode penelitian dilakukan dengan studi literatur terhadap sumber-sumber yang relevan, analisis pengetahuan terhadap parameter penentuan bahaya tsunami menggunakan algoritma *C 4.5*, dilakukan terhadap data sekunder penelitian Oktariadi. Penelitian ini mengimplementasi Algoritma *C 4.5* digunakan untuk menentukan tingkat bahaya tsunami di suatu daerah pesisir pantai mengacu pada kasus tingkat bahaya tsunami yang sudah ada. Dalam penelitian ini juga didapatkan hasil bahwa tingkat bahaya tsunami dibagi kedalam tiga klasifikasi yaitu tinggi, rendah dan sedang sebagai nilai dari atribut tujuan. Atribut sumber berupa landaan tsunami, keterangan pantai, kekasaran pantai dan intensitas gempa sebagai atribut sumber dengan nilai bervariasi sesuai kondisi wilayah. [4]

Saikhu, metode yang digunakan adalah *fuzzy decision tree*, yang memungkinkan dilakukannya prediksi suatu objek yang dimiliki lebih dari satu kelas. Dengan menerapkan teknik *data mining* pada diabetes yang diharapkan dapat ditemukan aturan klasifikasi untuk memprediksi potensi seseorang terserang diabetes, tanpa harus melawan diagnosis penyakit secara langsung dengan diagnosis ini. Dengan menggunakan 9 variabel yaitu *member of times pregnant (time)*, *plasma*

glucose concentration a 2 hours in an oral glucose tolerance test (OGIT), diastolic blood pressure (D), triceps skin fold thickness (T), 2-Hour serum Insulin (IPOST), body mass index (BMI), diabetes pedigree function (F), Age (T), dan Class variable (Diagnosa). Penelitian ini juga menerapkan algoritma C 4.5 pada data diabetes indian pima dengan hasil pembagian data himpunan *fuzzy* dengan menggunakan referensi data *standard* dengan *pruning* memiliki tingkat akurasi lebih tinggi dibanding dengan data kuartil. [5]

Ginting dkk, metode penelitian yang digunakan ini memiliki lima aspek utama yaitu, studi pustaka terhadap sumber-sumber yang mendukung penelitian ini, analisis Algoritma C 4.5 dalam penentuan masa studi mahasiswa, perancangan terhadap perangkat lunak (aplikasi) yang akan dibangun berdasarkan hasil yang diperoleh tahap analisis, implementasi terhadap perangkat lunak yang akan dibangun berdasarkan hasil yang diperoleh pada tahap perancangan, kemudian pengujian dan evaluasi terhadap hasil yang diperoleh berdasarkan kasus yang diteliti. Dalam penelitian ini juga didapatkan hasil bahwa jumlah *data training* mempengaruhi persentase kecocokan atau keakurasian. Persentase kecocokan hasil pengujian yang didapatkan dari *data training* 112 lebih kecil dibandingkan dengan *data training* 70 untuk berapapun kombinasi atribut. Demikian pula untuk semua konfigurasi atribut, persentase kecocokan hasil pengujian dengan *data training* 112 lebih kecil dibandingkan dengan *data training* 70. [6]

Jantan, metode penelitian yang digunakan ini memiliki empat aspek utama yaitu, *Managing Human Talent, Data Mining for Talent Management, Classification for Prediction, dan Decision Tree Techniques.* Dalam penelitian ini juga didapatkan

hasil bahwa penggunaan klasifikasi C 4.5 untuk memprediksi kinerja seorang karyawan untuk dipromosikan dari jabatannya memiliki potensi besar. Persentasi kecocokan hasil pengujian yang didapatkan dari 100 karyawan yang diuji adalah 77 persen karyawan layak untuk dipromosikan dan 23 persen karyawan tidak layak dipromosikan. [7]

Rahmi dkk, metode penelitian dilakukan dengan metode regresi karena metode regresi lebih baik digunakan dalam memprediksi jumlah produksi dibandingkan metode *fuzzy* madmani. Metode regresi menggunakan data historis untuk meramal dan mengasumsikan bahwa faktor yang diramalkan menunjukkan suatu hubungan sebab akibat dengan satu atau lebih variabel bebas, sehingga dari bentuk hubungan tersebut variabel tak bebas dapat digunakan untuk meramalkan nilai mendatang. Dalam penelitian ini juga didapatkan hasil bahwa prediksi harga saham berdasarkan data historis menggunakan algoritma genetika ini yaitu, representasi kromosom *real coded* yang digunakan mampu menyelesaikan permasalahan dalam memprediksi harga saham dengan ukuran populasi 1200, *crossover rate* (*cr*) 0,5 *mutation rate* (*mr*) 0,5 dan ukuran generasi 1500. [8]

Tabel 1.1 Tabel *State of The Art*

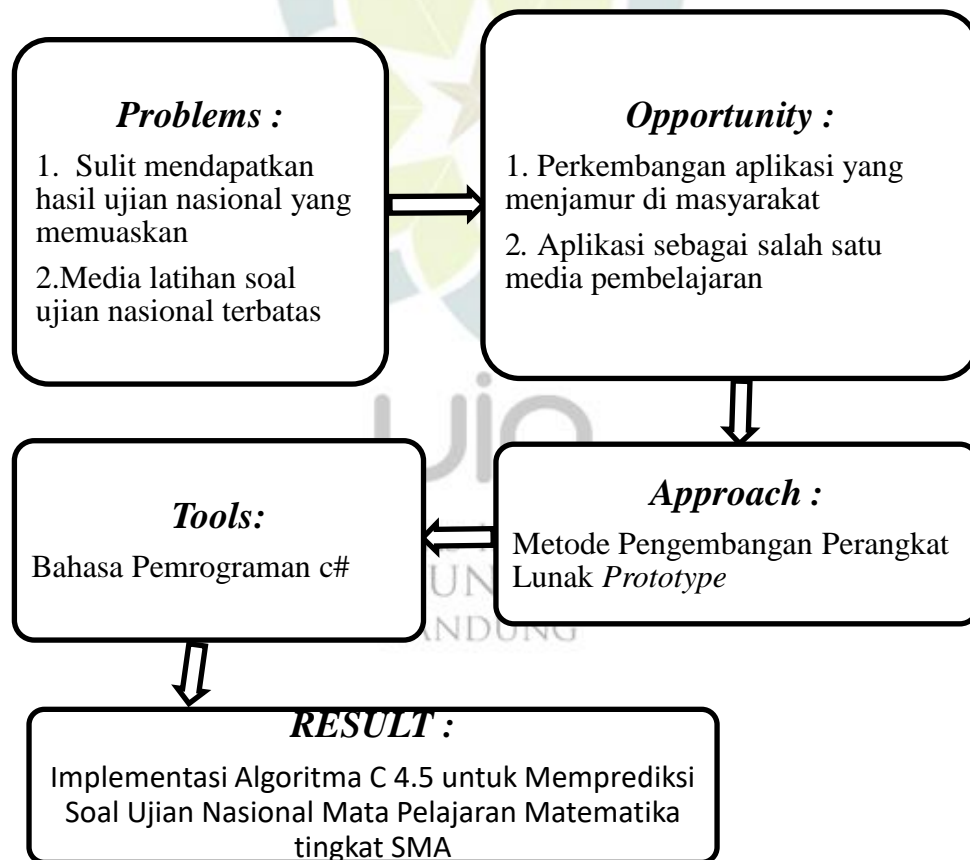
No.	Peneliti	Data	Metode	Teknologi
1.	Aa Zezen Zaenal Abidin (2011)	<ul style="list-style-type: none"> Indikator (Landaian Tsunami, Kelerangan Pantai, Kekasaran Pantai, Intensitas Gempa) Skala Kepekaan Terhadap Tsunami 	Algoritma C 4.5	Code Igniter, PHP

Tabel 2.1 Tabel State of The Art (lanjutan)

No.	Peneliti	Data	Metode	Teknologi
2.	Ahmad Saikhu (2011)	<ul style="list-style-type: none"> • <i>Member of times pregnant (time)</i> • <i>Plasma glucose concentration a 2 hours in an oral glucose tolerance test (OGIT)</i> • <i>Diastolic blood pressure (D)</i> • <i>Triceps skin fold thickness (T)</i> • <i>2-Hour serum Insulin (IPOST)</i> • <i>Body mass index (BMI)</i> • <i>Diabetes pedigree function (F)</i> • <i>Age (T)</i> • <i>Class variable (Diagnosa)</i> 	Algoritma C 4.5	Eclipse, Java
3.	Ginting, dkk (2014)	<ul style="list-style-type: none"> • Nilai mata kuliah (Algoritma Pemograman 1, Fisika Dasar 1, Fisika Dasar 2, Kalkulus 1, Kalkulus 2, Pengantar Sistem Komputer) • Masa Studi 	Algoritma C 4.5	Code Igniter, PHP
4.	Jantan (2010)	<ul style="list-style-type: none"> • Kategori • Jenis Kelamin • Kualifikasi • Lama Bekerja • Pengetahuann • Kualitas • Kontribusi • Evaluasi • Rekomendasi untuk promosi 	Algoritma C 4.5	Weka Machine Learning Software
5.	Rahmi dkk (2015)	<ul style="list-style-type: none"> • Populasi • Generasi • Periode • <i>Kromosom</i> • <i>Fitness</i> • <i>Crossover</i> • <i>Mutasi</i> 	Algoritma Genetika	MiniTab 16

1.6 Kerangka Pemikiran

Kerangka pemikiran adalah dasar pemikiran dari penelitian yang disintesis dari fakta-fakta, observasi dan telaah kepustakaan. Oleh karena itu kerangka pemikiran memuat teori atau konsep-konsep yang akan dijadikan dasar penelitian. Kerangka yang akan dijadikan sebuah pemikiran untuk proses penyusunan kajian dapat dilihat pada Gambar 1.1, yaitu :



Gambar 1.1 Kerangka Pemikiran

1.7 Metode Pengembangan Perangkat Lunak

1.7.1 Tahap pengumpulan data

Guna mendapatkan data yang diperlukan untuk membantu dalam penyusunan tugas akhir yang akan dibangun, maka digunakan metodologi sebagai berikut:

1. Teknik Wawancara

Teknik wawancara adalah proses memperoleh keterangan untuk tujuan penelitian dengan cara tanya jawab. Dalam penelitian ini yang menjadi sasaran wawancara yaitu para siswa SMA, guru-guru SMA serta Departemen Pendidikan Nasional maupun dengan teman guna mendapatkan informasi yang berkaitan dengan pokok bahasan.

2. Studi Kepustakaan

Teknik studi kepustakaan adalah proses memperoleh keterangan untuk tujuan penelitian dengan cara mencari data dalam buku-buku maupun jurnal-jurnal yang berhubungan dengan penelitian ini. Selain itu, studi kepustakaan juga dilakukan pada buku-buku latihan soal ujian nasional yang akan menjadi sumber materi yang akan diterapkan dalam aplikasi ini.

1.7.2 Metode Pengembangan Perangkat Lunak

Dalam pembangunan sistem ini mengikuti tahapan-tahapan berdasarkan metode yang digunakan yaitu *prototype*. Berikut tahapan-tahapan pembangunan sistem menggunakan metode *prototype* :

1. Pengumpulan kebutuhan

Pelanggan dan pengembang bersama-sama mendefinisikan format seluruh perangkat lunak, mengidentifikasi semua kebutuhan, dan garis besar sistem yang akan dibuat.

2. Membangun *prototyping*

Membangun *prototyping* dengan membuat perancangan sementara yang berfokus pada penyajian kepada pelanggan (misalnya dengan membuat input dan format output)

3. Evaluasi *protootyping*

Evaluasi ini dilakukan oleh pelanggan apakah *prototyping* yang sudah dibangun sudah sesuai dengan keinginan pelanggan. Jika sudah sesuai maka langkah 4 akan diambil. Jika tidak *prototyping* direvisi dengan mengulangi langkah 1, 2, dan 3.

4. Mengkodekan sistem

Dalam tahap ini *prototyping* yang sudah disepakati diterjemahkan ke dalam bahasa pemrograman yang sesuai.

5. Menguji sistem

Setelah sistem sudah menjadi suatu perangkat lunak yang siap pakai, harus dites dahulu sebelum digunakan. Pengujian ini dilakukan dengan White Box, Black Box, Basis Path, pengujian arsitektur dan lain-lain.

6. Evaluasi Sistem

Pelanggan mengevaluasi apakah sistem yang sudah jadi sudah sesuai dengan yang diharapkan. Jika ya, langkah 7 dilakukan; jika tidak, ulangi langkah 4 dan 5.

7. Menggunakan sistem

Perangkat lunak yang telah diuji dan diterima pelanggan siap untuk digunakan .

1.8 Sistematika Penulisan

Sistematika penulisan dari tiap bab dalam laporan tugas akhir ini bertujuan untuk mendapatkan keterarahan dan sistemasi dalam penulisan sehingga mudah dipahami, adapun sistematika secara umum dari penulisan laporan ini adalah:

BAB I : PENDAHULUAN

Bab I merupakan pengantar yang memberikan gambaran mengenai permasalahan- permasalahan yang kemudian akan dibahas pada bab-bab selanjutnya. Terdapat delapan pokok bahasan dalam bab ini, yaitu latar belakang, perumusan masalah, batasan masalah, tujuan, metodologi, *state of the art*, jadwal pembuatan dan sistematika penulisan.

BAB II : LANDASAN TEORI

Pada bab II akan dijelaskan tentang teori-teori yang digunakan dalam analisa permasalahan yang ada, dan juga teori-teori yang digunakan dalam perancangan dan implementasi.

BAB III : ANALISIS DAN PERANCANGAN

Bab III membahas mengenai analisis dari permasalahan yang ada saat ini dan analisis kebutuhan yang diperlukan untuk mengatasi permasalahan tersebut. Pembuatan desain dari sistem dengan mengacu pada analisis yang telah dibahas. Desain sistem yang akan dijelaskan terbagi menjadi tiga bagian, meliputi desain user interface, desain data dan desain proses.

BAB IV : IMPLEMENTASI DAN PENGUJIAN

Pada bab IV dijelaskan tentang spesifikasi aplikasi, kebutuhan aplikasi, implementasi aplikasi, dan pengujian yang dilakukan terhadap aplikasi yang dibangun.

BAB V : PENUTUP

Bab V berisi kesimpulan dan saran untuk pengembangan aplikasi lebih lanjut dalam upaya memperbaiki kelemahan pada aplikasi guna untuk mendapatkan hasil kinerja aplikasi yang lebih baik dan pengembangan program selanjutnya.

