

**KOMPARASI METODE KLASIFIKASI DATA MINING  
ALGORITMA C5.0 DAN NAÏVE BAYES UNTUK  
MEMPREDIKSI JURUSAN SISWA  
(STUDI KASUS: SMAN 1 GONDANGLEGI)**

**SKRIPSI**



**UNIVERSITAS ISLAM  
RADEN RAHMAT**

**MUHAMMAD ZAINURI  
NIM. 1857201001**

**UNIVERSITAS ISLAM  
RADEN RAHMAT**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM RADEN RAHMAT  
MALANG  
2022**

**KOMPARASI METODE KLASIFIKASI DATA MINING  
ALGORITMA C5.0 DAN NAÏVE BAYES UNTUK  
MEMPREDIKSI JURUSAN SISWA  
(STUDI KASUS: SMAN 1 GONDANGLEGI)**

**SKRIPSI**

**Diajukan kepada  
Universitas Islam Raden Rahmat**

**Untuk memenuhi salah satu persyaratan  
Dalam menyelesaikan program sarjana**



**MUHAMMAD ZAINURI** ★★★★★  
**NIM. 1857201001**

UNIVERSITAS ISLAM  
**RADEN RAHMAT**

**PROGRAM STUDI SISTEM INFORMASI  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM RADEN RAHMAT  
MALANG  
2022**

## LEMBAR PERSETUJUAN SKRIPSI

Judul : Komparasi Metode Klasifikasi Data Mining Algoritma  
C5.0 dan Naïve Bayes Untuk Memprediksi Jurusan Siswa  
(Studi Kasus: SMAN 1 Gondanglegi)

Penyusun : Muhammad Zainuri

NIM : 1857201001

Telah diperiksa dan disetujui untuk diuji pada tanggal 14 April 2022

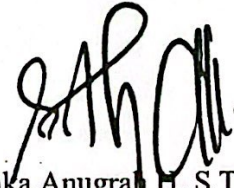
Disetujui oleh:

Pembimbing I



M. Hanif Fahmi, S.T., M.T  
NIDN.0710108403

Pembimbing II



Raka Anugrah H., S.TP., M.T  
NIDN.0728059101

## LEMBAR PENGESAHAN SKRIPSI

Judul : Komparasi Metode Klasifikasi Data Mining Algoritma C5.0 dan Naïve Bayes Untuk Memprediksi Jurusan Siswa (Studi Kasus: SMAN 1 Gondanglegi)  
Penyusun : Muhammad Zainuri  
NIM : 1857201001

Skripsi oleh Muhammad Zainuri ini telah dipertahankan di depan dewan penguji pada tanggal 14 April 2022

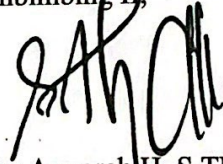
Disetujui oleh:

Pembimbing I,



M. Hanif Fahmi, S.T., M.T  
NIDN.0710108403

Pembimbing II,



Raka Anugrah H, S.TP., M.T  
NIDN.0728059101

Penguji I,



Farid Wahyudi, S.Kom, M.M  
NIDN.0705078805

Penguji II,



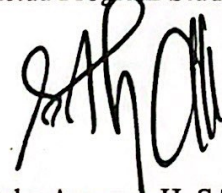
Novia Ratnasari, S.Pd., M.Pd  
NIY.2110500050

Mengesahkan,  
Dekan Fakultas Sains dan Teknologi



Mojib Rohman, S.Pd., M.Pd  
NIDN.0706088805

Mengetahui,  
Ketua Program Studi Sistem Informasi



Raka Anugrah H, S.TP., M.T  
NIDN.0728059101



## PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan dibawah ini:

Penyusun : Muhammad Zainuri

NIM : 1857201001

Program Studi : Sistem Informasi

Fakultas : Fakultas Sains dan Teknologi, Universitas Islam Raden  
Rahmat Malang

Menyatakan dengan sesungguhnya bahwa skripsi yang saya tulis ini benar-benar tulisan saya, dan bukan merupakan plagiasi baik sebagaimana seluruhnya.

Apabila dikemudian hari terbukti skripsi ini hasil plagiasi, baik sebagaimana atau seluruhnya, maka saya bersedia menerima sanksi atas perbuatan saya tersebut dengan ketentuan berlaku.

Malang, 25 April 2022  
Yang membuat pernyataan,



Muhammad Zainuri  
NIM. 1857201001

## ABSTRAK

**KOMPARASI METODE KLASIFIKASI DATA MINING ALGORITMA C5.0 DAN NAÏVE BAYES UNTUK MEMPREDIKSI JURUSAN SISWA (STUDI KASUS: SMAN 1 GONDANGLEGI)**

Muhammad Zainuri

NIM. 1857201001

Universitas Islam Raden Rahmat

Malang

2022

Penentuan jurusan siswa yang diimplementasikan di SMAN 1 Gondanglegi berdasarkan tes potensi akademik (TPA), tes *intelligence* (IQ) dan wawancara. Penjurusan tersebut dilakukan setelah siswa melakukan pendaftaran atau sebelum siswa diterima sebagai kelas X. Beberapa metode data mining dalam penentuan atau prediksi jurusan yang telah dilakukan oleh peneliti terdahulu diantaranya algoritma c4.5, c5.0 dan naïve bayes serta dalam perbandingan yaitu algoritma id3 dan c5.0 kemudian perbandingan naïve bayes dan decision tree. Yang mana dalam perbandingan antara metode algoritma c5.0 dan naïve bayes belum dilakukan. Dari permasalahan tersebut maka peneliti bertujuan untuk melakukan analisa perbandingan data mining menggunakan klasifikasi *algoritma c5.0* dan *naïve bayes* dalam memprediksi jurusan siswa. Adapun metode *data mining* yang digunakan yaitu *knowledge discovery in database* (KDD). Dalam proses KDD ada beberapa tahapan yakni Pemilihan data, Pembersihan data, Tranformasi data, *Data mining* dan Evaluasi. Berdasarkan hasil perbandingan pengujian yang telah dilakukan melalui berbagai skenario terhadap kedua metode tersebut, pengujian *10-fold cross validation* yang kemudian dicatat dalam confusion matrix menghasilkan nilai akurasi yaitu sebesar 60,87% untuk *algoritma c5.0* sedangkan untuk *naïve bayes* sebesar 56,52%. Dari hasil yang diperoleh *algoritma c5.0* merupakan metode paling baik dibanding *naïve bayes* yang dibuktikan dengan nilai tingkat akurasi yang dipaparkan lebih tinggi. Hasil akurasi tersebut didapatkan melalui software rapidminer.

Kata kunci: jurusan, *data mining*, *algoritma c5.0*, *naïve bayes*, *k-fold cross validation*, *confusion matrix*, akurasi, rapidminer

**ABSTRACT****DATA MINING ALGORITHM C5.0 AND NAÏVE BAYES  
CLASSIFICATION METHODS FOR PREDICTING STUDENT'S  
DEPARTMENT (CASE STUDY: SMAN 1 GONDANGLEGI)**

Muhammad Zainuri

NIM. 1857201001

Universitas Islam Raden Rahmat

Malang

2022

Determination of student majors implemented in SMAN 1 Gondanglegi based on academic potential tests (TPA), intelligence tests (IQ) and interviews. The assignment is done after the student registers or before the student is accepted as class X. Some data mining methods in determining or predicting majors that have been done by previous researchers include the algorithms c4.5, c5.0 and naïve bayes and in comparison, namely the id3 and c5.0 algorithms then the comparison of naïve bayes and decision trees. Which in comparison between the c5.0 algorithm method and naïve bayes has not been done. From these problems, researchers aim to analyze comparative data mining using the classification of c5.0 algorithms and naïve bayes in predicting student majors. The data mining method used is knowledge discovery in database (KDD). In the KDD process there are several stages, namely Data selection, data cleaning, data transformation, data mining and evaluation. Based on the results of comparisons of tests that have been conducted through various scenarios against the two methods, the 10-fold cross validation test which is then recorded in the confusion matrix produces an accuracy value of 60.87% for the c5.0 algorithm while for naïve bayes at 56.52%. From the results obtained by the c5.0 algorithm is the best method compared to naïve bayes as evidenced by the value of the accuracy level that is higher. The results of this accuracy are obtained through rapidminer software.

**Keywords:** major, data mining, c5.0 algorithm, naïve bayes, k-fold cross validation, confusion matrix, accuracy, rapidminer



## KATA PENGANTAR

Alhamdulillah puji syukur atas kehadiran Allah SWT yang telah melimpahkan Rahmat dan Hidayah-Nya, sehingga penulis dapat menyelesaikan tugas akhir skripsi ini dengan judul Komparasi Metode Klasifikasi Data Mining Algoritma C5.0 dan Naïve Bayes Untuk Memprediksi Jurusan Siswa (Studi Kasus: SMAN 1 Gondanglegi) sebagai persyaratan untuk lulus Sarjana (S1) Komputer di Universitas Islam Raden Rahmat Malang.

Penyelesaian skripsi ini tidak terlepas dari dukungan dan bantuan berbagai pihak. Oleh karena itu, atas nama pribadi penulis menyampaikan ucapan terima kasih kepada semua pihak yang telah membantu dalam penyusunan skripsi ini terutama kepada:

1. Kepada keluarga yang telah mendukung serta memotivasi saya untuk melakukan penelitian ini.
2. Bapak Mujibor Rohman, S.Pd., M.Pd selaku Dekan Fakultas Sains Dan Teknologi Universitas Islam Raden Rahmat Malang.
3. Bapak Raka Anugrah H, S.TP., M.T selaku Ketua Progam Studi Sistem Informasi dan dosen pembimbing II dalam penyusunan skripsi ini.
4. Bapak M. Hanif Fahmi, S.T., M.T selaku dosen pembimbing I dalam penyusunan skripsi ini.
5. Segenap dosen pengajar dan staf pada Program Studi Sistem Informasi Fakultas Saintek Unira Malang.
6. Bapak Fauzi Firman Alamsyah selaku Kepala IT SMAN I Gondanglegi.
7. Segenap Dewan Guru SMAN I Gondanglegi yang telah membantu dalam penelitian skripsi ini.
8. Seluruh teman yang telah membantu terutama rekan jurusan sistem informasi tahun angkatan 2018 dalam penyusunan skripsi ini yang tidak bisa disebutkan satu-persatu.
9. *Last but not least, I wanna thank me, for believing me, for doing all this hard work, for having no days off, for never quitting, for just being me at all times.*

Akhir kata, semoga segala bantuan yang telah diberikan oleh semua pihak di atas menjadi amalan yang bermanfaat. Dan dari penulisan skripsi yang penulis



susun, besar harapan tugas akhir skripsi ini bisa menjadi informasi yang bermanfaat bagi pembaca atau pihak lain yang membutuhkan.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dikarenakan terbatasnya pengalaman dan pengetahuan yang dimiliki penulis. Oleh karena itu, penulis mengharapkan segala bentuk saran serta masukan yang membangun agar penelitian menjadi lebih baik.

Malang, 26 Maret 2022

Penulis



UNIVERSITAS ISLAM  
**RADEN RAHMAT**

## DAFTAR ISI

	<b>Halaman</b>
LEMBAR JUDUL .....	i
LEMBAR PERSETUJUAN SKRIPSI.....	ii
LEMBAR PENGESAHAN SKRIPSI.....	iii
PERNYATAAN KEASLIAN TULISAN.....	iv
ABSTRAK .....	v
KATA PENGANTAR .....	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	4
1.3 Tujuan Penelitian.....	5
1.4 Batasan Masalah.....	5
1.5 Manfaat Penelitian.....	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 <i>Data Mining</i> .....	7
2.1.1 Pengelompokan <i>Data Mining</i> .....	8
2.1.2 Proses <i>Data Mining</i> .....	9
2.2 Klasifikasi.....	11
2.3 <i>Algoritma C5.0</i> .....	12
2.3.1 Pohon Keputusan.....	13
2.4 <i>Naïve Bayes</i> .....	14
2.5 <i>Cross Validation</i> .....	16
2.6 <i>Confusion Matrix</i> .....	16
2.7 Perangkat Lunak Pendukung.....	18
2.7.1 RapidMiner .....	18
2.7.2 Microsoft Exel .....	19
2.8 Jurusan Siswa .....	19
2.9 Penelitian Terdahulu.....	21
2.10 Kerangka Berfikir.....	24

BAB III METODE PENELITIAN.....	25
3.1 Kerangka Operasional .....	25
3.2 Tempat dan Waktu .....	26
3.3 Bahan dan Alat .....	27
3.4 Metode Penelitian.....	27
3.4.1 Teknik Pengumpulan Data.....	27
3.5 Metode Data Mining.....	28
3.5.1 Pengumpulan Data.....	28
3.5.2 <i>Data Selection</i> .....	28
3.5.3 <i>Data Cleaning (Pre-processing)</i> .....	28
3.5.4 <i>Data Transformation</i> .....	28
3.5.5 <i>Data Mining</i> .....	29
3.5.6 <i>Evaluation</i> .....	33
3.5.7 <i>Knowledge</i> .....	33
3.6 Teknik Analisis Data .....	33
BAB IV HASIL DAN PEMBAHASAN .....	35
4.1 Pengumpulan Data.....	35
4.2 Praproses Data .....	36
4.2.1 <i>Data Selection</i> .....	36
4.2.2 <i>Data Cleaning</i> .....	38
4.2.3 <i>Data Transformation</i> .....	41
4.3 Data Mining.....	43
4.3.1 Pembagian Data .....	43
4.3.2 Klasifikasi Metode <i>Algoritma C5.0</i> dan <i>Naïve Bayes</i> .....	46
4.5 Pengujian <i>Cross Validation</i> .....	48
4.6 Evaluasi Hasil Klasifikasi.....	51
4.6.1 Evaluasi Hasil Klasifikasi <i>Algoritma C5.0</i> .....	51
4.6.2 Evaluasi Hasil Klasifikasi <i>Naïve Bayes</i> .....	53
4.7 Perbandingan Klasifikasi <i>Algoritma C5.0</i> dan <i>Naïve Bayes</i> .....	54
4.8 Implementasi <i>Algoritma C5.0</i> dan <i>Naïve Bayes</i> .....	55
4.8.1 Implementasi <i>Algoritma C5.0</i> .....	55
4.8.2 Implementasi <i>Naïve Bayes</i> .....	58
4.9 Pembahasan .....	60
BAB V PENUTUP.....	64
5.1 Kesimpulan.....	64
5.3 Saran.....	64



DAFTAR PUSTAKA.....	65
LAMPIRAN.....	70



UNIVERSITAS ISLAM  
**RADEN RAHMAT**

## DAFTAR GAMBAR

	<b>Halaman</b>
Gambar 2.1 Proses Tahapan KDD (Fayyad,1996).....	9
Gambar 2.2 Kerangka Berfikir.....	24
Gambar 3.1 Kerangka Operasional.....	25
Gambar 3.2 Flowchart Algoritma C5.0.....	30
Gambar 3.3 Flowchart Naive Bayes .....	32
Gambar 4.1 Contoh Data Missing Value .....	39
Gambar 4.2 Proses Menghapus Missing Value .....	40
Gambar 4.3 Proses Pembagian Data .....	43
Gambar 4.4 Pembagian Data Latih 80% dan Data Uji 20% .....	43
Gambar 4.5 Hasil Pembagian Data 20% .....	44
Gambar 4.6 Hasil Pembagian Data 80% .....	44
Gambar 4.7 Proses Data Uji Decion Tree C5.0 .....	46
Gambar 4.8 Proses Data Uji Naive Bayes .....	46
Gambar 4.9 Hasil Akurasi Decision Tree C5.0.....	47
Gambar 4.10 Hasil Akurasi Naive Bayes .....	47
Gambar 4.11 Proses Cross Validation C5.0 dan NB .....	48
Gambar 4.12 0Proses K-fold Cross Validation NB .....	49
Gambar 4.13 Proses K-fold Cross Validation C5.0 .....	49
Gambar 4.14 Hasil Confusion Matrix Algoritma C5.0.....	51
Gambar 4.15 Hasil Akurasi, Precision, Recall.....	52
Gambar 4.16 Hasil Confusion Matrix Naive Bayes.....	53
Gambar 4.17 Hasil Akurasi, Precision, Recall.....	54
Gambar 4.18 Pohon Keputusan.....	57

**DAFTAR TABEL**

	<b>Halaman</b>
Tabel 2.1 Confucion Matrix.....	17
Tabel 2.2 confusion matrix 3 class.....	17
Tabel 2.3 Penelitian Terdahulu .....	21
Tabel 3.1 Confusion Matrix 3 Class .....	34
Tabel 4.1 Atribut Penelitian .....	35
Tabel 4.2 Contoh Data Sebelum Menentukan Class Target .....	36
Tabel 4.3 Contoh Data Setelah Menentukan Class Target .....	36
Tabel 4.4 Kategori Atribut MAT .....	37
Tabel 4.5 Kategori Atribut BIG .....	37
Tabel 4.6 Kategori Atribut BIN .....	37
Tabel 4.7 Kategori Atribut IPS .....	37
Tabel 4.8 Kategori Atribut IPA.....	37
Tabel 4.9 Kategori Atribut IQ.....	37
Tabel 4.10 Contoh Tabel Setelah Atribut Nama Dihapus.....	38
Tabel 4.11 Jumlah Data Missing Value .....	38
Tabel 4.12 Jumlah Data Missing Value Setelah Dihapus .....	40
Tabel 4.13 Jumlah Data Hasil Proses Pembersihan .....	40
Tabel 4.14 Pelabelan Atribut Jurusan .....	41
Tabel 4.15 Pelabelan Atribut MAT.....	41
Tabel 4.16 Pelabelan Atribut BIG.....	41
Tabel 4.17 Pelabelan Atribut BIN.....	42
Tabel 4.18 Pelabelan Atribut IPS.....	42
Tabel 4.19 Pelabelan Atribut IPA .....	42
Tabel 4.20 Pelabelan Atribut Score IQ .....	42
Tabel 4.21 Contoh Data Setelah Melakukan Transformasi .....	42
Tabel 4.22 Hasil Pembagian Data.....	45
Tabel 4.23 Contoh Data Latih.....	45
Tabel 4.24 Contoh Data Uji .....	45
Tabel 4.25 Akurasi .....	47
Tabel 4.26 Cross Validation Decision Tree C5.0 .....	50
Tabel 4.27 Cross Validation Naive Bayes .....	50
Tabel 4.28 Rumus Confusion Matrix 3 Class .....	51
Tabel 4.29 Hasil Performa klasifikasi Algoritma C5.0 dan Naive Bayes.....	54
Tabel 4.30 Hasil Nilai GainRatio.....	55



## DAFTAR LAMPIRAN

1. Contoh Data Penelitian
2. Hasil Pembagian (Data Latih)
3. Hasil Pembagian (Data Uji)



UNIVERSITAS ISLAM  
**RADEN RAHMAT**

## BAB I

### PENDAHULUAN

#### 1.1 Latar Belakang

Pendidikan adalah salah satu aspek yang penting dari kehidupan manusia. Perihal pentingnya pendidikan bagi masyarakat adalah untuk membentuk generasi cerdas berkualitas. Melalui proses pendidikan akan terbentuk pribadi yang unggul, berwawasan luas, bertanggung jawab serta mampu melewati tantangan di masa yang akan datang. Yusuf (2018) Pendidikan merupakan satu bidang yang berada di bawah tanggung jawab negara. Pembukaan UUD 1945 jelas mengamanatkan “mencerdaskan kehidupan bangsa”. Amanat tersebut tertanam secara terstruktur ke dalam berbagai undang-undang dan peraturan pendidikan. Undang-undang nomor 20 tahun 2003 tentang sistem pendidikan nasional disebutkan bahwa pendidikan adalah usaha sadar dan terencana untuk menciptakan suasana belajar dan proses pembelajaran bagi peserta didik untuk secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian, kecerdasan, akhlak mulia dan keterampilan yang berguna bagi dirinya, masyarakat, bangsa dan negara.

Bersamaan dengan perkembangan ilmu pengetahuan dan teknologi yang telah membawa perubahan dalam kehidupan manusia di mana dengan kemahiran dan peningkatan ilmu pengetahuan dan teknologi dapat memecahkan berbagai aspek permasalahan. Faizah dan Jananto (2021) menyatakan khususnya pada teknologi informasi penyimpanan data yang semakin kompleks memberikan banyak dampak terhadap perkembangan pengelolaan data dan penyimpanan data dalam jumlah besar (*big data*). Dengan semakin mudahnya menyimpan data dalam jumlah besar, tetap saja akan memberikan dampak menumpuknya data, sebagai contoh pada sebuah lembaga pendidikan. Kumpulan data sangatlah penting bagi lembaga pendidikan dalam hal evaluasi pendidikan. Atas adanya data besar yang dimiliki oleh lembaga pendidikan, maka dibutuhkan *data mining* dalam memenuhi kebutuhan informasi.

*Data mining* dalam penelitian Widaningsih (2019) merupakan prosedur yang menggunakan sistem statistika, matematika, kecerdasan buatan dan *machine learning* untuk mengekstrak dan mengenali informasi yang bermanfaat dari sebuah data dan pengetahuan yang terkait dari berbagai *database* besar. *Data mining* biasanya digunakan untuk mengelola data dengan ukuran besar dengan cara mencari pola yang sesuai dengan penerapan *data mining* tersebut. Hasil dari pengolahan *data mining* dapat digunakan untuk pengambilan keputusan maupun analisis yang diperlukan. Penjurusan menurut Gani (2013) dalam (Harianti, 2021) yaitu suatu prosedur penempatan dalam memilih program studi para siswa. Penjurusan ini merupakan proses penentuan atau memilih jurusan yang akan dilakukan oleh para siswa pada saat belajar di bangku SMA dan di perguruan tinggi. Penjurusan dilakukan agar para siswa dapat memilih jurusan sesuai dengan minat dan kemampuan akademik siswa.

SMAN 1 Gondanglegi adalah lembaga yang merupakan satu kesatuan dari Sekolah Menengah Atas (SMA) yang termasuk salah satu bentuk pendidikan formal, di mana sistem pendidikannya tidak hanya belajar mengenai mata pelajaran akan tetapi juga ada pembelajaran tambahan atau yang lebih dikenal dengan ekstrakurikuler. Sekolah menengah atas negeri (SMAN) 1 Gondanglegi berdiri sejak 22 Desember 1986 dengan terdapat tiga jurusan yaitu: IPA, IPS dan Bahasa. Kurikulum di SMAN I Gondanglegi yang digunakan adalah kurikulum 2013, di mana penjurusannya dimulai sejak kelas X (Sepuluh). Dalam mengelola sistem penjurusan ada beberapa pihak yang terlibat yaitu: waka kurikulum, tim guru yang bertugas serta guru bimbingan konseling. Sistem penjurusan yang diterapkan di SMAN 1 Gondanglegi adalah tes potensi akademik (TPA), tes *intelligence* (IQ) dan wawancara. Untuk penentuan jurusan pada SMA tersebut yang lebih dominan dilihat adalah dari nilai rata-rata tes potensi akademik siswa, yang mana pada saat wawancara siswa akan diberi rekomendasi oleh guru berdasarkan hasil dari nilai rata-rata tes potensi akademik. Soal yang dimuat dalam tes tersebut adalah mata pelajaran Matematika, Bahasa Inggris, Bahasa Indonesia, Ilmu Pengetahuan Alam dan Ilmu Pengetahuan Sosial. Dalam proses penjurusan yang telah dilakukan oleh SMAN 1 Gondanglegi terdapat kasus siswa yang meminta untuk pindah jurusan meskipun di awal penjurusan sudah diberitahu batas waktu tiga bulan tetapi ada



juga siswa yang meminta pindah jurusan lebih dari batas waktu yang telah diberikan pihak sekolah. Hal tersebut bisa dibilang bahwa sistem penjurusan yang digunakan di SMA tersebut kurang efektif dan kurang efisien, yang mana dalam pemilihan jurusan terdapat masalah baik dari siswa maupun guru.

Permasalahan dari siswa sendiri terjadi karena mereka kebingungan saat diperintahkan untuk memilih jurusan, maka dalam mengambil keputusan umumnya siswa meniru teman atau menuruti pilihan orang tuanya, sehingga terkadang membuat siswa tersebut berhenti ditengah jalan atau tidak melanjutkan sekolah karena jurusan yang diambil tidak sesuai dengan minat dan bakatnya. Sedangkan permasalahan dari guru adalah sebagaimana yang telah dijelaskan diatas yaitu guru lebih dominan melihat hanya pada nilai rata-rata tes potensi akademik siswa. Seharusnya pihak sekolah dalam menentukan jurusan harus melihat dari berbagai aspek yang relevan diantaranya dari segi minat siswa, siswa dapat diberi kesempatan untuk memilih satu jurusan yang menjadi minatnya. Yang mana tidak memperbolehkan siswa memilih lebih dari satu minat agar lebih mempermudah dalam proses penjurusan. Selain itu dapat juga melihat dari hasil nilai-nilai tes TPA yang disendirikan.

Untuk mengatasi permasalahan tersebut, dapat menggunakan *data mining* dengan model klasifikasi untuk menguji data siswa. Klasifikasi merupakan salah satu model *data mining* yang sering digunakan untuk menyelesaikan masalah penentuan dalam mengambil keputusan. Umam, dkk (2017) menyatakan bahwa klasifikasi adalah salah satu tugas penting dalam *data mining*. Tujuan utama dari klasifikasi yaitu untuk mengatur data dan mengelompokkannya ke dalam kelas yang berbeda. Definisi klasifikasi adalah fungsi yang membuat prediksi dan mengelompokkan data item tertentu dalam suatu kelas. Klasifikasi terdiri dari kumpulan data latih dengan kelas dan karakteristik yang telah ditentukan dan diketahui sebelumnya. Kinerja klasifikasi umumnya diukur dengan ketepatan atau akurasi. Beberapa model klasifikasi, diantaranya yaitu metode *decision tree*, *algoritma C4.5*, *naïve bayes*, *k-nearest neighbor*, *statistical analysis* dan *support vector machine*.

Hal ini selaras dengan penelitian terdahulu yang terkait antara lain: tentang Perbandingan *Algoritma ID3* dan *C5.0* Dalam Identifikasi Penjurusan Siswa SMA

oleh Munawaroh (2013). Dalam penelitian ini data yang digunakan sebanyak 200 data siswa yang dibagi menjadi 2 yaitu 150 *data training* dan 50 sebagai *data testing*, yang mana hasil nilai akurasi pada *algoritma C5.0* sebesar 95% dan *algoritma ID3* sebesar 93%. Hal ini membuktikan bahwa kinerja *algoritma C5.0* lebih baik dibandingkan *algoritma ID3*. Selanjutnya dalam penelitian Yulianti (2019) dengan judul Analisis Komparasi *Algoritma* Klasifikasi *Data Mining* Untuk Prediksi Penjurusan Siswa Menengah Atas (SMA) Pramita Karawaci Tangerang. Dalam penelitian ini data yang digunakan sebanyak 365 data siswa yang dibagi menjadi data training dan data testing dengan menguji data sebanyak 4 kali. Hasil nilai akurasi pada *naïve bayes* sangat baik yaitu sebesar 87,19% dibandingkan dengan decision tree sebesar 82,11%, sehingga *naïve bayes* merupakan metode yang cocok untuk diterapkan dalam menentukan jurusan siswa. Dari kedua penelitian yang telah dilakukan menunjukkan bahwa algoritma *c5.0* dan *naïve bayes* merupakan metode klasifikasi paling baik dengan hasil tingkat akurasi yang sangat tinggi yaitu sebesar 93% untuk algoritma *c5.0* dan 87,19% untuk *naïve bayes*.

Berdasarkan uraian permasalahan-permasalahan dalam latar belakang di atas, peneliti tertarik untuk membandingkan metode *algoritma c5.0* dan *naïve bayes* pada SMAN 1 Gondanglegi. Peneliti akan mengkomparasi metode klasifikasi *data mining algoritma c5.0* dan *naïve bayes* dalam memprediksi jurusan siswa untuk mendapatkan hasil tingkat akurasi paling tinggi, yang mana dapat digunakan sebagai solusi dari permasalahan yang ada di SMAN 1 Gondanglegi. Diharapkan penelitian ini akan memberi kemudahan dan dapat menjadi salah satu sarana untuk rekomendasi penjurusan siswa yang efisien dan efektif.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang penulis buat, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana tingkat akurasi antara metode *algoritma C5.0* dan *naïve bayes* untuk memprediksi jurusan siswa pada SMAN 1 Gondanglegi?
2. Manakah metode yang menghasilkan akurasi terbaik antara metode *algoritma C5.0* dan *naïve bayes* untuk memprediksi jurusan siswa pada SMAN 1 Gondanglegi?

### 1.3 Tujuan Penelitian

Berdasarkan rumusan yang penulis buat, maka tujuan dalam penelitian ini adalah:

1. Mengetahui tingkat akurasi antara *algoritma C5.0* dan *naïve bayes* untuk memprediksi jurusan siswa pada SMAN 1 Gondanglegi.
2. Melakukan perbandingan akurasi antara *algoritma C5.0* dan *naïve bayes* untuk memprediksi jurusan siswa pada SMAN 1 Gondanglegi.

### 1.4 Batasan Masalah

Agar masalah yang ada pada penelitian lebih terfokus, maka batasan masalah dalam penelitian adalah:

1. Data yang digunakan adalah data siswa kelas X (Sepuluh) jurusan IPA, IPS dan Bahasa di SMAN 1 Gondanglegi tahun ajaran 2021/2022.
2. Metode yang digunakan adalah komparasi antara *algoritma C5.0* dan *naïve bayes*.
3. Menggunakan Microsoft Excel untuk membantu dalam mengelola data dan *tool* Rapidminer untuk membantu dalam keakuratan hasil.
4. Pengujian dilakukan dengan teknik *cross validation* untuk memperoleh nilai akurasi yang maksimal.
5. Perbandingan akurasi dilakukan dengan menggunakan metode *confusion matrix* yang menghasilkan nilai *accuracy*, *precision*, *sensitifity*.

### 1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah:

1. Secara Teoritis

Penelitian ini dapat dijadikan sebagai bahan informasi data ilmiah guna mengembangkan ilmu pengetahuan pada lembaga pendidikan dan hasil dari penelitian dapat dijadikan sebagai pertimbangan agar dapat memberikan gambaran untuk peneliti selanjutnya dan untuk pihak lainnya yang membutuhkan dalam komparasi *data mining* terutama pada metode klasifikasi *algoritma C5.0* dan *naïve bayes*.



## 2. Secara Praktis

Bagi peneliti adalah dapat digunakan untuk menambah ilmu pengetahuan, melatih diri dalam menganalisa masalah serta memperoleh pemahaman yang berkaitan dengan *data mining*, khususnya pada metode klasifikasi *algoritma C5.0* dan *naïve bayes*. Sedangkan bagi tempat penelitian dapat menambah informasi terkait prediksi penjurusan dengan implementasi metode klasifikasi *algoritma C5.0* dan *naïve bayes*, yang mana bisa digunakan sebagai bahan pertimbangan pihak sekolah mengenai sistem penjurusan siswa yang lebih efektif dan efisien.



UNIVERSITAS ISLAM  
**RADEN RAHMAT**