

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan pada era 4.0 membawa perubahan yang begitu pesat dan berkembang hampir mencakup berbagai sektor. Perkembangan teknologi menghasilkan data yang besar dari berbagai bidang, baik dalam bidang industri, bidang ekonomi, bidang kesehatan, dan bidang lainnya, termasuk penerapan teknologi dalam bidang pendidikan yang dapat menghasilkan data yang cukup besar. Seiring dengan perkembangan teknologi, banyak sekali data yang dapat dimanfaatkan dan diolah menjadi sebuah informasi, termasuk teknologi informasi yang dapat membantu siswa dalam menentukan penjurusan pada tingkat Sekolah Menengah Atas (SMA) atau sederajat.

Saat ini sebagian besar siswa Sekolah Menengah Pertama (SMP) yang akan melanjutkan pendidikan ke jenjang selanjutnya yakni Sekolah Menengah Atas (SMA) cenderung belum mengetahui minat dan kemampuannya, sehingga siswa merasa kesulitan dalam memilih konsentrasi jurusan yang akan di pilih pada saat memasuki SMA. Pemilihan konsentrasi jurusan sangat penting karena mempengaruhi proses belajar siswa ke depannya, sehingga dalam memilih konsentrasi harus dipertimbangkan secara matang sesuai minat dan kemampuan akademik siswa. Beberapa kasus yang sering terjadi siswa memilih jurusan yang kurang tepat, sehingga menyebabkan siswa kesulitan dalam mengikuti dan memahami materi-materi yang diberikan, hal ini akan mempengaruhi proses belajar siswa yang tidak maksimal dan menurunnya prestasi akademik siswa karena siswa merasa kurang tepat dengan jurusan yang di pilihnya.

SMA Islam Ta'alumul Huda Bumiayu. Sekolah ini terletak di kabupaten Brebes bagian selatan beralamatkan di Jl. KH. Ahmad Dahlan no. 99 Desa Bumiayu Kecamatan Bumiayu Kabupaten Brebes. SMA Islam Ta'alumul Huda Bumiayu memiliki tiga (3) jurusan yaitu jurusan IPA, jurusan IPS dan jurusan Bahasa. Sistem penentuan jurusan sekolah ini merujuk pada peraturan Menteri Pendidikan dan Kebudayaan Republik Indonesia Nomor 81A tahun 2013 tentang Implementasi kurikulum mengemukakan bahwa penentuan jurusan pada sekolah

menengah atas, sekolah menengah kejurusan maupun madrasah Aliyah di lakukan langsung pada awal masuk sekolah atau pada saat siswa kelas sepuluh (X)[1]. SMA Islam Ta'alumul Huda Bumiayu merupakan salah satu SMA yang siswa siswinya cenderung banyak yang ragu dalam memilih konsentrasi jurusan yang akan di pilih.

Observasi dengan pihak SMA Islam Ta'alumul Huda mendapat informasi, SMA tersebut tidak menjadikan nilai ujian nasional sebagai acuan untuk menentukan jurusan bagi siswa. Sebelum menentukan jurusan, siswa/siswi wajib mengikuti *pretest* tertulis sebagai acuan untuk menentukan jurusannya. Hal ini membantu siswa yang masih ragu dalam menentukan jurusan mana yang akan di pilih, sehingga untuk meminimalisir memilih jurusan hanya mengikuti teman ataupun dorongan dari orang tua yang nantinya mengakibatkan siswa merasa salah memilih jurusan. Pada bagian kurikulum SMA Islam Ta'alumul Huda, masih kesulitan dalam mengelompokkan siswanya karena hasil *pretest* tidak sesuai dengan kriteria dari jurusan yang diminati siswanya.

Data mining merupakan proses yang menggunakan teknik statistik, matematika, kecerdasan buatan dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terakit dari berbagai *database* besar. Sementara itu, Data mining adalah suatu proses menemukan hubungan atau pola dari data yang tersimpan dalam *database* dengan teknik statistic atau matematika [2]

Penelitian sebelumnya yang dilakukan dengan judul “penggunaan teknik *unsupervised discretization* pada metode *Naïve Bayes* dalam menentukan jurusan siswa madrasah Aliyah”. Penelitian ini memprediksi jurusan siswa dengan memanfaatkan data nilai kumulatif siswa pada mata pelajaran pengetahuan alam, matematika, fisika, kimia dan biologi, serta nilai kumulatif mata pelajaran pengetahuan sosial, ekonomi, sejarah, geografi dan sosiologi. Penelitian ini menggunakan teknik *unsupervised discretization* dan metode *naïve bayes* untuk memprediksi jurusan siswa. Hasil pengujian pada 120 data sampel, penerapan teknik *unsupervised discretization* dan metode *naïve bayes* berhasil mengklasifikasi 111 data dan hasil presisi sebesar 92,8%.[3]

Kemudian pada penelitian yang lain dengan judul “prediksi pemilihan jurusan diperguruan tinggi”. Penelitian ini melakukan prediksi menggunakan algoritma *K-NN* dan *Naïve Bayes* dengan data sampel sebanyak 180 siswa. Data sampel terdiri dari *atribut* nilai semester 1 dan 2, mata pelajaran matematika, fisika, kimia, biologi, dan rata-rata semester. Kesimpulan dari hasil uji coba yang dilakukan, algoritma *Naïve Bayes* memiliki hasil akurasi yang lebih tinggi dengan hasil sebesar 90,34% daripada algoritma *K-NN* yang hasil akurasinya adalah 72,3%. [4]

Algoritma *K-Means* merupakan salah satu metode data *clustering non hierarchical* yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih *cluster*/kelompok. Metode ini mempartisi data ke dalam *cluster*/kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu *cluster* sesuai dengan karakteristiknya [5], Sedangkan algoritma *Naïve Bayes* (NB) adalah pengklasifikasi yang berdasarkan probabilitas bersyarat pada *Teorema Bayes*. NB dapat mengolah data dalam jumlah besar untuk memprediksi probabilitas di masa depan. NB juga memiliki beberapa kelebihan seperti mudah digunakan, hanya membutuhkan satu kali *scan* data *training*, penanganan nilai *atribut* yang hilang, dan data kontinyu. [6]. Penelitian sebelumnya, dapat disimpulkan bahwa algoritma *K-Means* dan NB bisa diterapkan pada penelitian ini yaitu memprediksi konsentrasi jurusan siswa SMA Islam Ta’alumul Huda Bumiayu karena mempunyai tingkat akurasi yang tinggi.

Data yang digunakan dalam penelitian ini merupakan data mentah, sehingga penelitian ini menggunakan dua algoritma yaitu algoritma *K-means* dan algoritma *Naïve Bayes*. Algoritma *K-Means* digunakan untuk mengelompokkan data sesuai dengan *Cluster* nya, sedangkan algoritma *Naïve Bayes* digunakan untuk memprediksi konsentrasi jurusan dengan memanfaatkan hasil keluaran algoritma *KMeans* sebagai label. Sehingga hasil prediksi ini dapat menjadi acuan bagi siswa dalam memilih konsentrasi jurusan. Maka dari itu, penulis mengangkat judul penerapan algoritma *K-Means* dan algoritma *Naïve Bayes* dalam pemilihan konsentrasi jurusan siswa di SMA Islam Ta’alumul Huda Bumiayu.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah pada penelitian ini adalah “berapa besar tingkat akurasi algoritma *K-Means* dan algoritma *Naïve Bayes* dalam memprediksi konsentrasi jurusan siswa di SMA Islam Ta’alumul Huda Bumiayu”.

1.3 Tujuan Penelitian

Sesuai dengan rumusan masalah diatas, penelitian ini bertujuan untuk mengetahui tingkat akurasi algoritma *K-means* dan algoritma *Naïve Bayes* dalam mengelompokkan dan memprediksi jurusan yang tepat untuk siswa dengan menerapkan metode data *mining*.

1.4 Manfaat Penelitian

Manfaat yang di harapkan dari penelitian ini yaitu:

1. Pihak instansi
 - a. Membantu sekolah dalam mengelompokkan jurusan untuk siswa sesuai dengan kemampuan akademiknya.
 - b. Membantu sekolah sebagai bahan pertimbangan dalam menentukan jurusan untuk siswa.
 - c. Memberikan acuan atau aturan yang mempermudah sekolah dalam mengelompokkan penjurusan siswa.
2. Pihak siswa
 - a. Membantu siswa dalam memilih jurusan yang diminati dan sesuai dengan kemampuan akademiknya.
 - b. Memberikan acuan bagi siswa dalam memilih konsentrasi jurusannya.
3. Pihak peneliti
 - a. Peneliti dapat memahami konsep data mining algoritma *K-means Clustering*, dan algoritma *Naïve Bayes*.
 - b. Peneliti dapat menerapkan algoritma *K-means Clustering* dan algoritma *Naïve Bayes* dalam mengelompokkan dan memprediksi konsentrasi jurusan siswa SMA Islam Ta’alumul Huda Bumiayu.

1.5 Ruang Lingkup

Berdasarkan identifikasi masalah yang telah diuraikan diatas, maka penulis membatasi ruang lingkup pada :

1. Data yang digunakan adalah data siswa SMA Islam Ta'alumul Huda Bumiayu tahun pelajaran 2019/2020.
2. Algoritma yang digunakan adalah Algoritma *K-means Clustering* dan *Naïve bayes*
3. Parameter yang digunakan adalah data nilai *pretest* IPA, IPS, Basaha Indonesia, dan Matematika.
4. Bahasa pemrograman yang digunakan *PHP* dan *MySQL*.

1.6 Sistematika Penulisan

Sistematika penulisan terdiri dari 5 bab, yaitu :

BAB I PENDAHULUAN

Pada bab ini berisi latar belakang, manfaat dan tujuan, rumusan masalah, batas masalah, sistematika penulisan

BAB II TINJAUAN PUSTAKA

Dalam bab ini membahas tentang dasar teori yang mendukung pembahasan penulisan ini

BAB III METODELOGI PENELITIAN

Bab ini menjelaskan tentang kerangka pemikiran penerapan metode ini dan apa saja yang dibutuhkan dalam perancangan ini.

BAB IV HASIL PENELITIAN DAN PEMBAHASAN

Pada baba ini pembahasan dari hasil eksperimen , evaluasi dan validasi yang dilakukan.

BAB V PENUTUP

Bab penutup berisi kesimpulan hasil penelitian dan saran peneliti selanjutnya.