

BAB 3

METODE PENELITIAN

Tujuan utama dari penelitian ini adalah untuk mengetahui minat bakat calon mahasiswa berdasarkan tes minat bakat yang dilakukan pada “BakatMu System” dan juga untuk mengetahui metode machine learning yang tepat untuk studi kasus minat bakat ini. Pada bagian ini akan menjelaskan tentang hasil penelitian yang telah dilakukan seperti informasi dataset yang digunakan, framework sistem, dan perancangan sistem.

3.1 LOKASI PENELITIAN

Lokasi pengambilan *sampel* data yaitu menggunakan publik dataset¹.

Table 3. 1 Dataset 1

Dataset Name	Field	Total
Dataset-1	Linguistic Verbal Logika Matematika Spasial Visual Ritmik Musik Kinestetik Interpersonal Intrapersonal Naturalis Eksistensial	1.443

¹ <https://gitlab.com/ekadadang12/datasetminatbakat>

3.1.1 Data Jenis Kecerdasan

Pada penelitian ini daftar nama jenis kecerdasan tersebut akan diberi nomor urut otomatis, menggunakan kode “K001” untuk urutan pertama, “K002” untuk urutan kedua dan seterusnya. Berikut data jenis kecerdasan:

Table 3. 2 Data Jenis Kecerdasan

Kode	Jenis Kecerdasan	Jurusan
K001	Kecerdasan Linguistic-Verbal	Ilmu Perpustakaan Ilmu Komunikasi Bahasa dan Sastra Ilmu Hubungan Internasional Ilmu Hukum Ilmu Politik
K002	Kecerdasan Logika–Matematika	Statistika Administrasi Negara Akuntansi Ilmu Ekonomi Pendidikan Matematika Ilmu Fisika Ilmu Kimia Teknik Informatika Sistem Informasi
K003	Kecerdasan Spasial-Visual	Seni Rupa Teknik Arsitektur Planologi Teknik Sipil
K004	Kecerdasan Ritmik-Musik	Seni Musik
K005	Kecerdasan Kinestetik	Kedokteran Gigi Kebidanan Seni tari PJKR Teknik Mesin
K006	Kecerdasan Interpersonal	Ilmu sosiologi

		PGPAUD PGSD Psikologi Kedokteran Ilmu Keperawatan Fakultas Kesehatan Masyarakat Antropologi
K007	Kecerdasan Intrapersonal	Ilmu Agama Administrasi Niaga
K008	Kecerdasan Naturalis	Kedokteran Hewan Fakultas Perikanan Fakultas Peternakan Ilmu Biologi Fakultas Pertanian
K009	Kecerdasan Eksistensial	Ilmu Filsafat Ilmu Sejarah

3.1.2 Skoring

Skoring digunakan untuk mengukur kecerdasan berdasarkan kepada 9 kategori dalam kecerdasan. Skor disesuaikan dengan pilihan jawaban. Berikut ini penjelasan dari skor:

1. Skor 4 diberikan, jika memilih jawaban setuju.
2. Skor 3 diberikan, jika memilih jawaban kurang setuju.
3. Skor 2 diberikan, jika memilih jawaban tidak setuju.
4. Skor 1 diberikan, jika memilih jawaban sangat tidak setuju. Dengan pemberian skor pada masing-masing pertanyaan akan diperoleh skor akhir, yaitu dengan cara menjumlahkan skor dari tiap-tiap kecerdasan lalu diambil skor yang paling tinggi untuk mengetahui kecerdasan yang lebih dominan. Jika ada persamaan skor maka diambil keduanya.

3.2 SARANA PENDUKUNG

Dalam penelitian ini, digunakan publik dataset untuk mendukung evaluasi tes minat bakat. Dataset merupakan kumpulan *datatable* yang memiliki relasi antar *dataTable* (*data relation*) atau di dalam satu database bisa terdapat banyak *data Table* yang berelasi. Di Dalam dataset ini terdapat objek yang merepresentasikan data dan relasinya di memori.

3.3 TEKNIK PENGUMPULAN DATA

Teknik pengumpulan data dilakukan untuk mencari data dan mengumpulkan data agar dapat diolah menjadi informasi yang bermanfaat. Pada penelitian ini metode pengumpulan data yang digunakan adalah:

1. Observasi

Pada metode ini dilakukan pengamatan dengan memberikan pertanyaan atau kuesioner kepada calon mahasiswa untuk mendapatkan informasi yang berhubungan dengan topik tugas akhir.

2. Studi Pustaka

Selain menggunakan metode observasi, penelitian ini juga mengambil data dari buku, jurnal, *e-book* serta sumber-sumber lainnya seperti laman web, artikel dan dokumendokumen yang berkaitan dengan topik.

3.4 UJI COBA INSTRUMEN

Untuk mengetahui data yang dikumpulkan adalah data yang valid dan reliabel maka diperlukan uji coba instrumen. Uji coba instrumen ini dimaksudkan untuk mengetahui validitas dan reliabilitas instrumen, sehingga setelah data teruji validitas dan reliabilitasnya baru dapat digunakan. Langkah-langkahnya adalah sebagai berikut:

- a. Uji Validitas

Instrument penelitian sebelum dipakai untuk pengambilan data maka perlu di uji cobakan. Uji coba dengan mengambil sampel sebanyak 156 calon mahasiswa. Butir pertanyaan pada instrument penelitian dikatakan valid, jika setelah diuji menggunakan statistik nilai r hitungnya (pearson correlation) lebih besar dari r tabel, sedangkan jika nilai r hitung lebih kecil dari r tabel maka butir pertanyaan tersebut tidak valid atau gugur. Adapun besarnya nilai r tabel dapat dicari berdasarkan jumlah sampel dan taraf signifikansinya. Berdasarkan hasil uji coba diperoleh butir soal yang gugur adalah nomor 6. Jumlah butir soal yang valid adalah 107 (untuk hasil secara lengkap dapat dilihat pada lampiran 1).

- b. Uji Reliabilitas

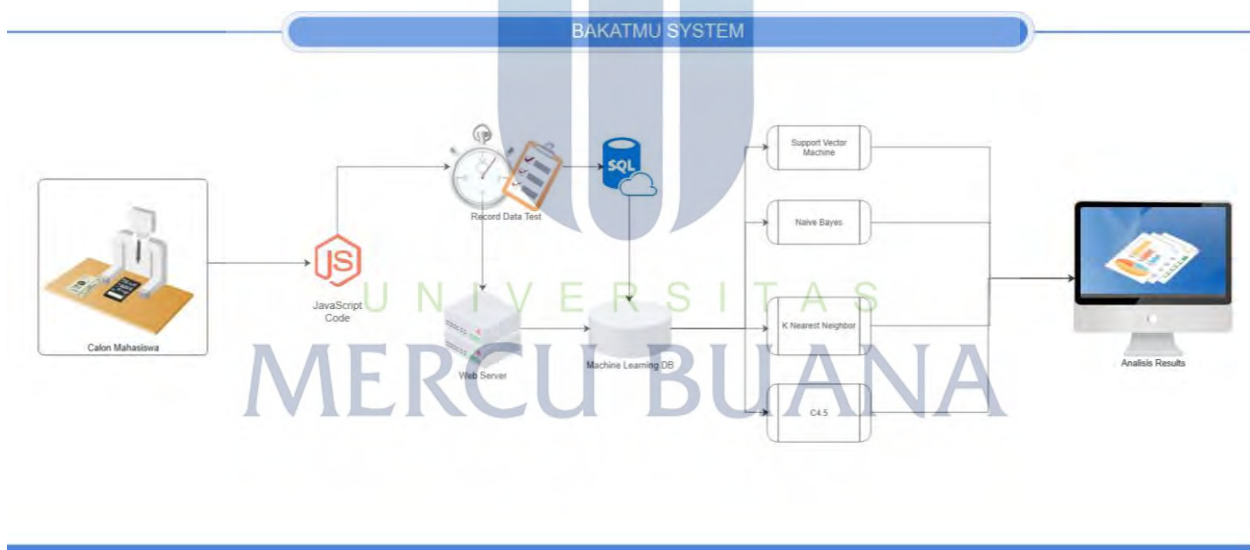
Dalam penelitian ini untuk mencari realibilitas instrumen digunakan rumus Alpha Cronbach. Kriteria pengujian instrumen dikatakan handal apabila r hitung lebih besar dari pada r tabel pada taraf signifikan. Untuk hasil uji reliabilitas secara ringkas analisisnya adalah seperti pada tabel 3.3.

Table 3. 3 Hasil Uji Reabilitas Cronbach Alfa

Variabel	Realibilitas	Interpretasi
Minat Bakat Calon Mahasiswa	0.97	Sangat Reliabel

3.5 FRAMEWORK SISTEM

Sistem ini memiliki 2 user yaitu admin dan calon mahasiswa, admin memiliki akses untuk menginputkan data soal, kategori jurusan, jurusan, kategori kecerdasan, kampus, karir, dan mapping kecerdasan. Sedangkan calon mahasiswa memiliki akses untuk melakukan tes dan menginputkan jawaban dari tes yang dilakukan. Inputan dari jawaban tes yang dilakukan oleh calon mahasiswa di proses oleh sistem sehingga akan memberikan hasil berupa jurusan yang sesuai dengan bakat dan minat calon mahasiswa tersebut. Sedangkan *output* yang akan didapatkan oleh admin berupa dashboard data analytic yang berisi *Accuracy* masing-masing *classifiers*. Adapun framework dari penelitian ini dapat dilihat pada Gambar 3.1



Gambar 3. 1 Framework Sistem

Pada framework sistem di atas digambarkan bahwa user melakukan tes minat bakat pada sistem dengan cara mengisi soal-soal yang tersedia pada sistem, kemudian setelah user menginput jawaban, maka sistem akan melakukan *preprocessing* dan menyimpan data kedalam database. Setelah itu akan dilakukan proses *classifiers* menggunakan metode *machine learning*. Kemudian masuk pada tahap *feature representation* untuk pemberian label pada dataset. Pada proses ini metode digunakan diantaranya *Naïve Bayes*, *K-Nearest Neighbor*, dan *Support Vector Machine* yang akan menghasil *positive* dan *negative value*

sehingga bisa menghasilkan prediksi jurusan yang sesuai dengan minat bakat user yang mengikuti test.

Skenario Workflow:

1. Inputan Data

Pada proses ini calon mahasiswa melakukan tes minat bakat pada sistem dengan cara mengisi soal-soal yang tersedia pada *bakatmu system*. Soal-soal yang diberikan terdiri dari:

Table 3. 4 Soal Tes Minat Bakat

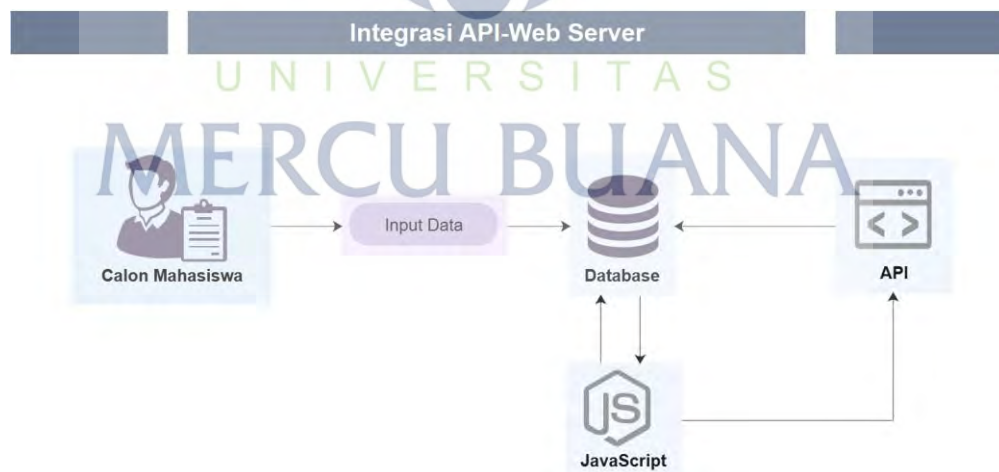
No	Soal
1	Suka bercerita.
2	Memiliki ingatan yang baik untuk hal-hal yang sepele.
3	Menyukai permainan kata-kata.
4	Hobi membaca buku.
5	Seorang pembicara yang baik.
6	Menyukai pelajaran bahasa daripada matematika dan ilmu alam.
7	Senang membicarakan dan menulis ide-ide yang ada di otak.
8	Memiliki kosakata yang cukup banyak sehingga tidak ada hambatan atau kesulitan dalam berkomunikasi dengan berbagai kalangan.
9	Senang menulis karangan, atau segala hal yang berbentuk tulisan.
10	Lebih memilih untuk menulis dalam presentasi kelompok.
11	...
100	Relaksasi dan meditasi latihan yang bermanfaat.
101	Suka mengunjungi situs alam.
102	Menikmati membaca filsuf kuno dan modern.
103	Belajar hal-hal baru lebih mudah ketika memahaminya.
104	Ingin mengetahui apakah ada bentuk-bentuk lain dari kehidupan di alam semesta.
105	Mempelajari sejarah dan budaya kuno membantu memberi suatu pandangan yang nyata.
106	Hobi membaca buku sejarah.
107	Cukup Fasih dalam menjelaskan budaya kuno.
108	Mempunyai minat cukup besar dalam mempelajari sejarah.

2. Record Data Test

Setelah dilakukan input data oleh calon mahasiswa maka system akan melakukan record data tes minat bakat menggunakan javascript code dan kemudian akan mengirimkan data pada web server untuk dilakukan *preprocessing* dan menyimpan data kedalam database. Adapun proses dari *pre-processing* tersebut adalah seperti, data hasil test yang diinputkan calon mahasiswa akan di convert ke dalam bentuk dataset agar bisa diolah menggunakan pemodelan *machine learning*, kemudian dataset akan diproses menggunakan empat metode machine learning yaitu C4.5, *K-Nearest Neighbor*, Naïve Bayes dan *Support Vector Machine*. Pada proses pemodelan ini akan didapatkan *accuracy* dan prediksi dari masing-masing model. *Accuracy* dari keempat model *machine learning* ini akan di bandingkan dan kemudian akan diambil prediksi dari model yang memiliki *accuracy* tertinggi. Hasil dari pengolahan data ini berupa *endpoint API*. Dimana API ini nantinya akan diintegrasikan dengan bakatmu system.

3. JavaScript (JS) dan Web Server

Sistem akan mengirimkan data-data yang telah disubmit oleh calon mahasiswa ke JS terlebih dahulu kemudian pada JS, terdapat proses pemanggilan API untuk dilakukan pengiriman data pada web server. Adapun proses pemanggilan API terlihat seperti gambar 3.2.



Gambar 3.2 Integrasi API – Webservice

Gambar diatas merupakan proses integrasi API- Web Server pada Bakatmu system, dimana data yang diinputkan mahasiswa akan tersimpan pada database, kemudian data tersebut akan diolah dan akan menghasilkan sebuah API, API tersebut akan di initiate pada JS untuk bisa diintegrasikan dengan proses backend system.

4. *Machine Learning DB*

Machine Learning DB merupakan database yang digunakan untuk menyimpan semua data data Bakatmu Sistem. Pada proses sebelumnya, JS akan melakukan pemanggilan API untuk melakukan penyimpanan data pada database. Jika proses penyimpanan data tersebut berhasil dilakukan, maka data-data test yang sudah dilakukan akan tersimpan ke dalam database.

Setelah data-data tersimpan pada database, proses selanjutnya adalah proses classifiers menggunakan metode machine learning. Kemudian masuk pada proses *feature representation* untuk pemberian label pada dataset. Pada proses ini metode yang digunakan diantaranya Naïve Bayes, *K-Nearest Neighbor*, dan *C4.5* yang akan menghasilkan *positive* dan *negative value* sehingga bisa menghasilkan prediksi jurusan yang sesuai dengan minat bakat user yang mengikuti test. Untuk proses klasifikasi ini akan dilakukan menggunakan python yang akan menghasilkan sebuah endpoint *Application Programming Interface (API)*. Kemudian API tersebut akan dipanggil kembali pada *services* bakatmu system. Proses metode *machine learning* ini akan dilakukan pada database dengan proses sebagai berikut:

a. Support Vector Machine

Support Vector Machine (SVM) dikenalkan pertama kali oleh Vapnik tahun 1992 sebagai salah satu metode ML yang bekerja dengan prinsip Structural Risk Minimization (SRM) yang bertujuan untuk menemukan hyperplane terbaik yang memisahkan dua buah class pada input space. Metode ini menggunakan hipotesis berupa fungsi-fungsi linear dalam sebuah ruang fitur yang berdimensi tinggi, dengan mengimplementasikan learning bisa berasal dari teori pembelajaran statistic. Tingkat akurasi ada pada model yang akan dihasilkan oleh proses peralihan dengan SVM sangat bergantung terhadap fungsi kernel dan parameter yang digunakan.

Data pada suatu dataset diberikan variabel x_i , sedangkan untuk kelas pada dataset diberikan variabel y_1 . Metode SVM membagi dataset menjadi 2 kelas. Kelas pertama yang dipisah oleh hyperplane bernilai 1, sedangkan kelas lainnya bernilai -1 [8].

$$[(w^T \cdot x_i) + b] \geq 1 \text{ untuk } y_i = +1 \quad (1)$$

$$[(w^T \cdot x_i) + b] \leq -1 \text{ untuk } y_i = -1 \quad (2)$$

dimana, X_i mendeskripsikan mengenai data ke $-i$, W menjelaskan mengenai nilai bobot support vektor yang tegak lurus dengan hyperplane, B merupakan nilai bias, dan Y_i mempresentasikan mengenai kelas data ke $-i$.

b. *Naïve Bayes*

Algoritma ini merupakan sebuah metode klasifikasi menggunakan metode probabilitas dan statistic, dimana algoritma ini memprediksi peluang yang akan terjadi dimasa yang akan datang berdasarkan pengalaman di masa sebelumnya. Prediksi bayes ini memiliki rumus seperti persamaan berikut[5]:

$$P(H|E) = \frac{P(E|H) \times p(H)}{p(E)} \tag{3}$$

Penjelasan dari persamaan diatas adalah $P(H|E)$ didefinisikan sebagai probabilitas akhir (*Posterior*) dimana sebuah hipotesis H terjadi jika diberikan bukti (*evidence*) E terjadi, $P(E|H)$ merupakan probabilitas suatu bukti E terjadi maka mempengaruhi hipotesis H (*Likelihood*), $P(H)$ didefinisikan sebagai probabilitas awal (*Prior*) hipotesis H terjadi tanpa melihat bukti apapun, dan $P(E)$ merupakan probabilitas awal bukti E yang terjadi tanpa melihat hipotesis/bukti yang lainnya.

Pada metode Bayes ini, hasil dari hipotesis atau peristiwa (H) bisa diperkirakan berdasarkan pada beberapa bukti (E) yang diamati.

c. *K Nearest Neighbor*

Algoritma ini melakukan klasifikasi proyeksi data berdimensi banyak. Proses klasifikasi dilakukan dengan mencari titik c terdekat dari c -baru (*nearest neighbor*). Teknik pencarian tetangga terdekat yang umum dilakukan dengan menggunakan formula jarak euclidean.

Metode ini akan mencari pola ruang K data training yang paling dekat. K data training ini adalah k nearest neighbor dari unknown sample “Closeness: dinyatakan dengan euclidean distance, dimana euclidean distance di antara dua titik,

$$X = x_1, x_2, \dots, \dots, x_n \tag{4}$$

Dan

$$Y = y_1, y_2, \dots, \dots, y_n” \tag{5}$$

Untuk menghitung *Euclidean distance* atau menghitung jarak dalam *K-Nearest Neighbor* digunakan fungsi berikut ini :

$$euc = \sqrt{\sum_{i=1}^n (x_{2i} - x_{1i})^2} \quad (6)$$

Dimana, X_2 merupakan data latih, X_1 merupakan data uji, i merupakan variable data, dan n merupakan dimensi data.

Untuk melakukan pembobotan variabel, maka dilakukan dengan metode perbandingan pasangan, bobot variabel ditentukan dengan normalisasi vektor *eigen* yang diasosiasikan dengan nilai eigen pada suatu matriks rasio sebelum membandingkan harus ditentukan terlebih dahulu skala nilai pengaruh atau penting antara variable[6].

5. Analisis Results

Pada tahapan analisis results ini, system akan memberikan hasil analisa yang telah dilakukan system berupa rekomendasi jurusan berdasarkan hasil test yang disubmit calon mahasiswa.

