

BAB 2

LANDASAN TEORI

Pada bagian ini kami akan menjelaskan mengenai beberapa penelitian terkait yang menggunakan pemodelan machine learning Naïve Bayes, K-Nearest Neighbor, C4.5 dan Support Vector Machine.

2.1 Pengantar Minat Bakat

Minat merupakan suatu sikap yang dilakukan secara terus-menerus dan mampu membuat pola perhatian seseorang sehingga dalam menjadi selektif terhadap objek minatnya. Sedangkan bakat diartikan sebagai kemampuan bawaan atau potensi diri yang masih perlu dikembangkan dan dilatih. Bakat memerlukan latihan dan pendidikan agar bermanfaat dimasa yang akan datang. Kemampuan ini menunjukkan bahwa suatu tindakan yang dapat dilakukan pada saat ini. Sedangkan prestasi merupakan perwujudan dari bakat dan kemampuan (Dita Merawati, 2019).

Menurut Wita Yulianti (2016), minat merupakan masalah yang sangat penting didalam sebuah pendidikan. Karena melalui tes minat dan bakat, calon mahasiswa dapat mengetahui potensi yang dimiliki dirinya termasuk kelebihan dan kekurangannya, baik dari segi akademis maupun kepribadian.

Salah satu cara untuk mengetahui minat dan bakat dari seseorang adalah dengan melakukan serangkaian tes minat dan bakat. Tes minat dan bakat ini dapat digunakan untuk memetakan seseorang dalam pemilihan bidang yang sesuai dengan kompetensi dan keterampilan yang dimiliki. Tujuan dari tes minat bakat ini adalah agar seseorang dapat meningkatkan kelebihan dan mengatasi kelemahan yang dimiliki sehingga individu tersebut dapat mengembangkan potensi yang dimiliki (Dita Merawati, 2019).

Pada penelitian sebelumnya sudah terdapat beberapa metode yang digunakan dalam melakukan analisis tes minat bakat ini, salah satunya adalah dengan menggunakan metode machine learning. Proses dari penerapan metode tersebut adalah dengan melakukan data selection, preprocessing, transformation, data mining dan evaluation. Namun masih terdapat beberapa kekurangan seperti banyaknya waktu yang diperlukan dalam proses klasifikasi data dan besarnya jumlah memori yang dibutuhkan. Untuk itu, dalam penelitian ini, kami mencoba untuk melakukan perbandingan beberapa metode machine learning technique dalam melakukan sebuah analisis terhadap tes minat bakat.

(Dita Merawati, 2019) telah melakukan penelitian tes minat bakat dengan menggunakan metode C4.5, dimana pada penelitian ini analisa metode C4.5 menggunakan tools yang sudah ada dan untuk tes minat bakat sendiri masih dilakukan secara manual belum menerapkan metode C4.5 di dalam sistem tersebut. Pada penelitian ini juga belum memberikan rekomendasi jurusan untuk calon mahasiswa.

Irham Cahya Nugraha, dkk (2016), telah melakukan penelitian tes minat bakat dengan menggunakan sistem pakar, namun pada penelitian ini hanya memberikan kesimpulan kecerdasan yang dimiliki calon mahasiswa, tidak memberikan rekomendasi jurusan yang tepat untuk calon mahasiswa.

2.2 Artificial Intelligence

Kecerdasan buatan atau *artificial intelligence* (AI) adalah teknik yang digunakan untuk menyelesaikan sebuah masalah dengan menirukan kecerdasan yang dimiliki oleh makhluk hidup. Terdapat tiga metode yang dikembangkan untuk membuat AI, diantaranya (M. Marlina et al, 2017):

- a. *Fuzzy Logic* (FL). Teknik ini biasanya digunakan oleh mesin untuk beradaptasi dengan kebiasaan makhluk hidup untuk menentukan sebuah keputusan yang tidak kaku 0 atau 1. Salah satu penerapan logika *fuzzy* ini adalah untuk sistem pengereman kereta api di Jepang.
- b. *Evolutionary Computing* (EC). Pendekatan ini menggunakan skema evolusi jumlah individu yang banyak untuk menyeleksi individu terbaik guna membangkitkan generasi selanjutnya.
- c. *Machine Learning* (ML) atau pembelajaran mesin merupakan teknik yang paling banyak digunakan untuk menirukan perilaku manusia dalam menyelesaikan masalah.

Menurut M.Sunaryo et al (2020), sistem pakar merupakan salah satu bagian dari kecerdasan buatan (Artificial Intelligence) yang khusus untuk melakukan pemecahan masalah dan memberikan solusi di tingkat ahli. Konsep dasar dari sistem pakar adalah keahlian yang ditransfer ke suatu komputer, keahlian yang berupa pengetahuan tersebut disimpan dan nantinya dapat digunakan oleh sistem agar mencari solusi dari fakta-fakta yang didapatkan. Tujuan utama dari sistem pakar merupakan sebagai penghubung antara seorang pakar dengan user yang membutuhkan pengetahuan di bidang kepakaran tersebut.

Sistem pakar merupakan berbasis komputer yang digunakan untuk penyelesaian masalah yang hanya bisa diselesaikan oleh ahli/pakar yang ahli pada bidang tertentu. Dengan menggunakan sistem ini, maka masyarakat dapat melakukan perhitungan seperti seorang

pakar (Kusumadewi, 2003). Pakar yang dimaksud disini merupakan orang yang memiliki keahlian khusus yang dapat menyelesaikan masalah-masalah yang tidak bisa diselesaikan oleh masyarakat umum. Keuntungan dari sistem pakar ini adalah (M. Marlina et al, 2017):

1. Membuat seorang yang tidak memiliki keahlian tertentu, bisa menjadi seperti layaknya seorang yang memiliki keahlian dibidang tersebut
2. Dapat bekerja dengan informasi yang tidak lengkap atau tidak pasti
3. Meningkatkan *output* dan produktivitas sehingga mengurangi jumlah pekerja yang dibutuhkan, dan akhirnya mereduksi biaya
4. Meningkatkan kualitas
5. Sistem pakar dapat menghasilkan *output* yang konsisten dan mengurangi kesalahan
6. Sistem pakar dapat melatih pekerja yang tidak berpengalaman karena sudah menggunakan peralatan yang kompleks dan mudah dioperasikan
7. Handal (*reliability*)
8. Sistem pakar memiliki kemampuan untuk memecahkan masalah yang kompleks
9. Sistem pakar merupakan arsip yang terpercaya dari keahlian seorang pakar, sehingga user seolah-olah berkonsultasi langsung dengan pakar.

M. Marlina et al (2017) mengatakan bahwa *machine Learning* (ML) atau pembelajaran mesin merupakan pendekatan AI yang digunakan untuk menirukan perilaku manusia untuk menyelesaikan masalah atau melakukan otomatisasi. Dalam ML terdapat dua aplikasi utama yaitu klasifikasi dan prediksi. Klasifikasi adalah sebuah metode yang terdapat pada ML untuk memilah atau mengklasifikasikan objek berdasarkan ciri tertentu. Sedangkan prediksi atau regresi digunakan untuk menerka *output* dari suatu data dan *input* berdasarkan data *training*. Dimana ciri khas dari ML ini adalah adanya proses pelatihan, pembelajaran dan *training*.

Machine learning memiliki tiga kategori yaitu supervised leaning, unsupervised learning, reinforcement learning (Somvanshi & Chavan, 2016). Dimana teknik yang digunakan oleh supervised learning adalah metode klasifikasi kumpulan data sepenuhnya diberikan label untuk dilakukan klasifikasi kelas yang tidak dikenal. Unsupervised learning atau yang sering dikenal sebagai cluster tidak membutuhkan pemberian label dalam kumpulan data sebelumnya dan hasilnya tidak mengidentifikasi label yang telah ditentukan (Thupae et al , 2018). Sedangkan reinforcement learning menurut Board (2017) biasanya berada diantara supervised learning dan unsupervised learning, teknik yang digunakan bekerja dalam lingkungan yang dinamis dan menyelesaikan tujuannya tanpa adanya pemberitahuan secara eksplisit dari computer jika tujuan tersebut tercapai (Das & Nene, 2017).

Pada penelitian kali ini, kategori yang digunakan adalah supervised learning. Dimana pada supervised learning memiliki beberapa algoritma diantaranya Naïve bayes, K-Nearest Neighbor, C4.5 dan Support Vector Machine.

2.2.1 Naïve Bayes Classifier

Naïve Bayes merupakan salah satu teknik prediksi berbasis *probabilistic* sederhana yang berdasarkan pada penerapan teorema Bayes (aturan bayes) dengan asumsi independensi yang kuat.

Model yang digunakan pada metode ini adalah “model fitur independen”. Dimana dalam bayes, makna dari independen yang kuat pada fitur adalah sebuah fitur dalam suatu data tidak berkaitan dengan ada atau tidaknya fitur lainnya dalam data yang sama. Naïve Bayes Classifier ini dapat diatih dengan efisien dalam teknik supervised learning, dimana proses yang dilakukan menganalogikan ada atau tidaknya suatu variable pada suatu kelas yang tidak memiliki korelasi dengan atau tanpa variable lain dikelas yang sama (Muhammad, 2017).

Theopilus Bayu Sasongko et al, (2019) telah melakukan penelitian dengan menggunakan naïve bayes classifiers ini, namun pada penelitiannya menyatakan terdapat kelemahan pada algoritma naïve bayes classifiers yaitu ketika menjalankan proses training dengan menggunakan dataset yang memiliki bobot attribute tidak seragam, maka akan menyebabkan performance klasifikasi dari metode ini tidak sesuai. Edy Widodo, (2020) pada penelitiannya mengatakan bahwa metode naïve bayes memiliki hasil akurasi yang akurat karena semakin banyak data yang digunakan maka semakin baik dan akurat hasil yang diperoleh.

Pada penelitian lainnya dengan studi kasus yang sama juga mengimplementasikan algoritma naïve bayes. Namun pada penelitian tersebut mengalami permasalahan mengenai probabilitas kondisional dapat bernilai nol dan nilai nol tersebut dianggap tidak berlaku dalam nilai probabilitas prediksi (Kencana & Astuti, 2016). Hairani et al, (2020) juga mengimplementasikan metode naïve bayes dalam penelitiannya untuk klasifikasi jurusan pada sebuah sekolah menengah atas, pada penelitiannya memperoleh akurasi naïve bayes sebesar 99,16%, sensitivitas 99,58%. Dengan demikian dalam penelitiannya mengusulkan untuk menggunakan metode naïve bayes untuk studi kasus klasifikasi jurusan.

Penelitian selanjutnya yang menggunakan metode naïve bayes dengan studi kasus yang sama dilakukan oleh Agung jasuma et al, (2019) pada penelitian tersebut mengatakan bahwa potensi diri yang dimiliki berperan penting terhadap perolehan IPK masing-masing mahasiswa. Penelitian tersebut berhasil memperoleh akurasi sebesar 90,65% dalam melakukan klasifikasi mahasiswa berdasarkan potensi diri.

2.2.2 K-Nearest Neighbor

Nearest Neighbor Classifiers berdasarkan pada proses pembelajaran dengan analogi atau *learning by analogi*. Dimana data training dideskripsikan dalam bentuk atribut numerik n-dimensi. Setiap sample akan mewakili sebuah titik pada ruang n-dimensi. Maka semua data training disimpan pada pola ruang n-dimensi. Ketika diberikan “*unknown*” sample, maka metode ini akan mencari pola ruang K data *training* yang paling dekat. Untuk melakukan pembobotan variable, 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 (M.A. Yaqin , 2019).

Penelitian terkait yang menggunakan metode KNN dilakukan oleh Vergy ayu kusumadewi et al, (2020) mengenai klasifikasi jurusan siswa dengan menggunakan 288 data. Pada penelitiannya berhasil memperoleh akurasi sebesar 85,15% dengan nilai $k = 5$. Selanjutnya penelitian mengenai penentuan bidang berdasarkan minat bakat menggunakan metode KNN yang berhasil memperoleh akurasi sebesar 78,91% dengan responden sebanyak 1376 santri yang berusia 13-15 tahun. Pada penelitian ini KNN tidak lebih unggul dibandingkan beberapa metode lainnya (Moh. Jasri et al, 2017).

Moh. Ainol Yaqin, (2017) juga melakukan penelitian terkait studi kasus minat bakat menggunakan metode KNN, pada penelitian tersebut memperoleh akurasi KNN sebesar 54,61% dari 10 k-fold validasi dengan $k = 5$. Penelitian lainnya oleh Abdul Rahman kadafi yang menjelaskan tentang klasifikasi penjurusan pada salah satu sekolah menengah atas, pada penelitian tersebut metode KNN memperoleh akurasi sebesar 77,86% dengan perolehan area under curve sebesar 0,5 (Abdul Rahman kadafi, 2018). Penelitian selanjutnya yaitu dilakukan oleh Harafani, (2018) yang menyatakan bahwa KNN memiliki beberapa kelemahan salah satunya adalah pengaruh nilai K terhadap hasil akurasi dan prediksi yang diperoleh. Pada penelitian tersebut metode KNN menghasilkan akurasi sebesar 82% dengan nilai $k = 1$.

Berdasarkan penjelasan penelitian terkait diatas dapat dilihat keoptimalan nilai k pada metode KNN sangat berpengaruh terhadap akurasi dan hasil prediksi yang diperoleh.

2.2.3 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 (M.Jasri, 2017).

Beberapa penelitian terkait yang menggunakan metode support vector machine yaitu seperti yang dilakukan pada penelitian Edy Widodo, (2020) mengenai prediksi jurusan mengatakan bahwa metode SVM memiliki hasil yang tidak stabil apabila semakin banyak data yang digunakan serta banyaknya waktu yang diperlukan dalam proses klasifikasi data dan besarnya jumlah memori yang dibutuhkan dan pada penelitian tersebut akurasi yang dihasilkan belum memenuhi ketepatan pengklasifikasian data. Suyanto pada penelitiannya mengatakan bahwa SVM memiliki beberapa kelemahan diantaranya yaitu sulit untuk diaplikasikan pada data yang memiliki jumlah sample dan dimensi yang sangat besar, selain itu pada umumnya metode SVM ini diformulasikan untuk menyelesaikan masalah klasifikasi dua kelas, walaupun bisa digunakan untuk klasifikasi multiclass tapi performance dari metode ini tidak bisa maksimal. (Suyanto, 2017).

Penelitian lainnya mengenai prediksi pengambilan mata kuliah menggunakan metode SVM dengan hasil pengujian menggunakan sample 3 mata kuliah. Pada penelitian tersebut melakukan perbandingan antara metode SVM dengan KNN dan hasil penelitiannya mengatakan bahwa SVM tidak lebih unggul dari metode KNN karena KNN memiliki kinerja yang jauh lebih baik dan stabil dalam melakukan klasifikasi mahasiswa pengambil mata kuliah (Fitra A. Bachtiar et al, 2018). Penelitian selanjutnya juga dilakukan oleh Muhammad Ichwan et al, (2018) dengan studi kasus yang berbeda menggunakan metode SVM dan menggunakan metode KNN sebagai metode pembandingan pada penelitian tersebut. Pada penelitian ini menggunakan 24 object pengujian dan mendapatkan tingkat akurasi untuk metode SVM sebesar 87.5% namun pada penelitian ini juga menyebutkan kelemahan dari SVM yang sulit untuk diaplikasikan pada himpunan data dalam jumlah besar karena menyebabkan ketidakstabilan pada hasil klasifikasi dan akurasi. Selanjutnya penelitian untuk menentukan kelulusan mahasiswa menggunakan metode SVM menghasilkan nilai probabilitas sebesar 0,898 sehingga penelitian tersebut menyimpulkan bahwa SVM memiliki kemampuan untuk menggeneralisasi masalah walaupun memiliki sample terbatas, namun terdapat kekurangan dari

metode SVM yaitu jumlah atribut dalam data dapat mempengaruhi bahkan mengurangi tingkat akurasi dan menambah kompleksitas dari metode tersebut (Rizky Agung Permana, 2016).

2.2.4 Metode C4.5

Metode C4.5 adalah struktur *flowchart* yang mirip seperti struktur pohon, setiap titik pohon merupakan yang telah uji, dan setiap cabang merupakan hasil uji, titik akhir merupakan pembagian kelas yang dihasilkan. Pohon keputusan merupakan metode klasifikasi dan prediksi yang sangat kuat dan banyak digunakan. Metode ini dapat mengubah fakta yang sangat besar sehingga menjadi pohon keputusan yang merepresentasikan aturan. Aturan tersebut dapat dengan mudah dipahami dan dapat diekspresikan dalam bentuk basis data seperti SQL (*Structured Query Language*) untuk mencari record pada kategori tertentu (Supena et al, 2018).

Moh Royandi Azkia et al, (2018) pernah menggunakan metode C4.5 pada penelitiannya untuk pemilihan program studi calon peserta didik. Dimana pada penelitiannya berhasil mendapatkan tingkat akurasi sebesar 68% setelah melakukan pengujian dengan menggunakan K-Fold Cross Validation dengan nilai K=10. Menurutnya metode ini bisa digunakan untuk mengatasi studi kasus tersebut. Beti Novianti et al, (2016) juga pernah menggunakan metode ini dalam penelitiannya untuk penjurusan siswa SMA dan penelitian tersebut berhasil mendapatkan akurasi sebesar 89,74%. Selain itu, penelitian lainnya yang menerapkan algoritma C4.5 ini dilakukan oleh Dewi Anggraeni et al, (2019) yang menyatakan bahwa penelitian yang dilakukan untuk penentuan kesesuaian jurusan mahasiswa berhasil mendapatkan tingkat keakuratan sebesar 93,31% dan akurasi rekomendasi jurusan sebesar 82,64%.

Penelitian selanjutnya yang menerapkan metode C4.5 dilakukan oleh Afif Fauzi et al, (2020) untuk prediksi pemilihan jurusan pada salah satu universitas. Berdasarkan pengujian dengan menggunakan confusion matrix terhadap 158 data nilai calon mahasiswa, berhasil memperoleh tingkat akurasi sebesar 87,50% dengan predikat good classification. Selain itu pada penelitian tersebut juga mengatakan bahwa metode C4.5 memiliki performa yang lebih baik dibandingkan dengan metode lainnya.

Annisa Ulfa et al, (2020) juga menerapkan metode C4.5 pada penelitiannya untuk menentukan rekomendasi jurusan kuliah bagi calon mahasiswa, pada penelitiannya menghasilkan akurasi hanya sebesar 12% dengan tingkat error 80%. Hal ini disebabkan karena data yang digunakan tidak sesuai dalam merekomendasikan jurusan kuliah menggunakan metode ini.

2.3 Penelitian Terkait

Berikut ini merupakan ringkasan dari daftar penelitian yang terkait dengan tes minat bakat dengan menggunakan *machine learning*.

Table 2. 1 Literature Review 1

No	Sumber	Masalah dan Tujuan	Metode	Hasil
1.	Penerapan Data Mining Penentu Minat Dan Bakat Siswa Smk Dengan Metode C4 . 5, <i>J. Algor</i> , vol. 1, no. 1, pp. 28–37, 2019	Untuk mempermudah penentuan hasil minat dan bakat seseorang setelah melakukan tes secara manual.	C4.5	Pada penelitian ini didapatkan bahwa metode C4.5 dapat digunakan sebagai salah satu metode pada kasus penentuan minat dan bakat siswa SMK tingkat keakuratan yang didapatkan pada tes bakat mencapai 82,65% dan pada tes minat mencapai 90,23%.
2.	Implementasi Metode K-Nearest Neighbor untuk Rekomendasi Peminatan Studi (Studi Kasus: Jurusan Teknik Informatika Universitas Brawijaya), <i>Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer</i> , vol. 2, no. 7, pp. 2745–2753, 2018.	Mahasiswa diharapkan dapat memilih keminatan yang sesuai untuk studinya maka di perlukan sistem rekomendasi untuk keminatan mahasiswa dengan harapan mahasiswa dapat memilih keminatan yang sesuai dengan bakat, keinginan serta tentu saja nilai mata kuliah wajib yang sudah ditempuh. Dalam penelitian ini digunakan metode KNN karena metode	KNN	Berdasarkan pengujian yang telah dilakukan dari 30 data uji dengan membandingkan antara hasil sistem dengan data yang diperoleh dari pakar maka diperoleh hasil akurasi sebesar 76,66%

		<p>ini adalah metode yang efektif apabila mempunyai data latih yang besar.</p>		
--	--	--	--	--



UNIVERSITAS
MERCU BUANA

3.	<p>Penerapan Metode Naïve Bayes Classifier Untuk Penjurusan Siswa Pada Madrasah Aliyah Al-Falah Jakarta, Fountain of Informatics Journal, vol. 5, no. 1, pp. 2548–5113, 2020.</p>	<p>Pihak sekolah Madrasah Aliyah Jakarta dalam kesulitan mengklasifikasikan siswa berdasarkan hasil tes penjurusan. Saat ini teknik penjurusan yang dilakukan sekolah hanya mengandalkan pengolahan data dan pengurutan dengan Microsoft Excel. Sistem klasifikasi penjurusan siswa yang dikembangkan dapat membantu proses penjurusan siswa dengan lebih mudah, cepat, dan akurat.</p>	<p>Naïve Bayes</p>	<p>Berdasarkan hasil pengujian dan uji coba terhadap metode dan prototype yang telah dikembangkan diperoleh kesimpulan bahwa metode Naïve Bayes dapat digunakan untuk mengklasifikasikan jurusan siswa. Namun demikian, proses pengubahan nilai numerik menjadi nilai kategorikal menyebabkan hasil akurasi, presisi, dan recall menjadi rendah.</p>
----	---	---	--------------------	--