BAB II. TINJAUAN PUSTAKA

Bab II digunakan untuk menjelaskan penelitian terdahulu dan teori-teori yang berkaitan dan mendukung penelitian.

2.1. Studi Sebelumnya

Perkembangan penelitian mengenai penggunaan *Artificial Intelligence* (AI) berbasis *chatbot* dalam pembelajaran serta keterkaitan gaya belajar dengan interaksi digital telah berkembang pesat dalam beberapa tahun terakhir. Berbagai penelitian telah membuktikan pengaruh *chatbot* terhadap perilaku dan hasil belajar mahasiswa. Namun, sebagian besar studi masih berfokus pada efektivitas penggunaan AI secara umum dan belum secara spesifik mengkaji keterkaitan antara gaya belajar dan pola *prompting* dalam interaksi dengan *chatbot*. Berikut beberapa penelitian relevan yang menjadi landasan dalam penelitian ini.

Pratiwi, dkk. (2024) [9]menganalisis persepsi mahasiswa Indonesia terhadap penggunaan *ChatGPT* dalam pembelajaran Bahasa Indonesia menggunakan pendekatan survei. Hasil penelitian menunjukkan bahwa mahasiswa memandang *ChatGPT* sebagai alat bantu yang efektif untuk memahami materi dan menyusun tulisan akademik. Namun, ditemukan variasi dalam cara mahasiswa memanfaatkan *ChatGPT*, sebagian menggunakannya secara pasif untuk mendapatkan jawaban langsung, sementara sebagian lain menggunakannya dengan cara mencoba-coba dan terus memperbaiki pertanyaannya. Temuan ini mengindikasikan adanya perbedaan individual dalam pola interaksi dengan AI yang diduga berkaitan dengan preferensi belajar masing-masing mahasiswa, meskipun penelitian ini belum secara spesifik menghubungkan pola tersebut dengan model gaya belajar tertentu.

Fauzi, dkk. [10] mengkaji peran *ChatGPT* dalam meningkatkan produktivitas belajar mahasiswa di perguruan tinggi melalui pendekatan kualitatif berbasis wawancara mendalam. Hasil penelitian menunjukkan bahwa mahasiswa yang menggunakan *ChatGPT* secara terarah dan terstruktur mengalami peningkatan produktivitas yang lebih signifikan dibandingkan penggunaan yang dibandingkan yang menggunakannya

tanpa arah yang jelas. Penelitian ini menemukan pola pemanfaatan yang beragam di antara mahasiswa, di mana strategi interaksi dengan chatbot mencerminkan cara berpikir yang berbeda-beda. Meskipun demikian, penelitian ini belum mengaitkan perbedaan pola interaksi tersebut dengan dimensi gaya belajar secara formal.

Oviedo Ramirez, dkk. [11]melakukan *systematic literature review* untuk mengeksplorasi hubungan antara gaya belajar dan sumber belajar digital dalam konteks *Adaptive Learning Systems* (ALS). Hasil kajian menunjukkan bahwa *Felder-Silverman Learning Styles Model* (FSLSM) merupakan model gaya belajar yang paling banyak digunakan dalam pengembangan ALS. Setiap dimensi FSLSM punya ciri khas cara belajar yang bisa dikaitkan dengan jenis sumber belajar digital tertentu, sehingga memberi dasar yang kuat bagi perancang sistem untuk menghubungkan sumber digital dengan profil gaya belajar pengguna demi mendukung pembelajaran yang lebih sesuai dengan kebutuhan masing-masing orang. Penelitian ini menjadi dasar pemikiran utama dalam penelitian ini karena menyediakan kerangka yang menghubungkan gaya belajar dengan pola interaksi digital, namun belum mengeksplorasi pola interaksi berbasis chatbot berbentuk percakapan sebagai cerminan gaya belajar.

Sawalha, dkk. [7]menganalisis pola *prompting* mahasiswa dan pengaruhnya terhadap performa *ChatGPT* menggunakan pendekatan *mixed methods* dan *thematic analysis*. Hasil penelitian menunjukkan bahwa mahasiswa mengembangkan kecenderungan *prompting* individual yang berbeda-beda, mencerminkan cara berpikir yang khas dalam memecahkan masalah. Kualitas dan struktur prompt secara langsung memengaruhi seberapa tepat dan seberapa dalam jawaban *ChatGPT*, di mana pola *prompting* yang lebih rinci dan terus diperbaiki menghasilkan *output* yang lebih sesuai dengan kebutuhan belajar mahasiswa. Penelitian ini menjadi salah satu dasar pemikiran penting karena membuktikan berdasarkan data nyata bahwa pola *prompting* bukan sekadar perbedaan teknis, melainkan mencerminkan cara berpikir masing-masing individu, meskipun belum menghubungkannya secara langsung dengan model gaya belajar FSLSM[12].

Hacer Güner dan Erkan Er [8]mengeksplorasi interaksi mahasiswa dengan *ChatGPT* dalam pembelajaran pemrograman menggunakan pendekatan eksperimen yang tidak sepenuhnya terkontrol dan analisis profil. Hasil penelitian mengidentifikasi lima profil interaksi mahasiswa dengan *ChatGPT*, yaitu *AI-Reliant Code Generators*, *AI-*

Reliant Code Generator & Refiners, *AI-Collaborative Coders*, *AI-Assisted Code Refiners*, dan *AI-Independent Coders*. Profil-profil ini mencerminkan perbedaan kemandirian dan cara berpikir mahasiswa dalam memanfaatkan AI, yang secara tidak langsung menunjukkan adanya hubungan antara pendekatan belajar masing-masing individu dengan pola interaksi yang mereka gunakan. Meskipun penelitian ini tidak menggunakan kerangka FSLSM secara langsung, temuan ini mendukung asumsi penelitian ini bahwa pola *prompting* dapat menjadi cerminan dari karakteristik gaya belajar mahasiswa.

Ayyoub dan Al-Kadi [13]mengembangkan pendekatan identifikasi gaya belajar secara otomatis menggunakan *semi-supervised machine learning* berbasis data aktivitas digital mahasiswa. Hasil penelitian menunjukkan bahwa gaya belajar mahasiswa berhasil dikenali secara otomatis dari data aktivitas digital dengan hasil yang cukup baik tanpa perlu banyak data yang diberi label secara manual. Penelitian ini membuktikan bahwa pola aktivitas digital memiliki potensi yang kuat sebagai sinyal tidak langsung untuk mengenali karakteristik gaya belajar individu[14]. Temuan ini menjadi dasar penting bagi penelitian ini dalam mengeksplorasi pola *prompting chatbot* sebagai representasi karakteristik gaya belajar mahasiswa, meskipun penelitian tersebut menggunakan data LMS konvensional dan belum mengeksplorasi data interaksi *chatbot conversational* sebagai sumber identifikasi.

Pritalia, dkk. [15] mengembangkan sistem klasifikasi gaya belajar *Sequential-Global* dalam pembelajaran multimedia menggunakan data *eye tracking* dan *machine learning*. Penelitian ini mengumpulkan data rekaman gerakan mata dari 68 partisipan menggunakan *eye tracker* GP3 selama mereka mengamati stimulus visual dalam konteks pembelajaran. Fitur-fitur diekstrak dari data *eye tracking* kemudian diklasifikasikan menggunakan *algoritme Support Vector Machine, KNN, Logistic Regression, dan Naive Bayes dengan ground truth label* dari kuesioner ILS. Hasil terbaik diperoleh dengan *Support Vector Machine* yang mencapai akurasi 71%. Penelitian ini menjadi fondasi penting karena membuktikan bahwa perilaku digital yang dapat direkam dalam hal ini gerakan mata berpotensi mengidentifikasi gaya belajar FSLSM secara otomatis, meskipun pendekatan berbasis fitur yang digabungkan masih menghasilkan akurasi terbatas.

Nurlatifa, dkk. [16]melanjutkan penelitian Pritalia, dkk. (2020) dengan mengembangkan kerangka klasifikasi yang lebih canggih menggunakan pendekatan *time series* dan *deep learning*. Dengan dataset yang sama (68 partisipan), data eye tracking diolah sebagai *time series* berdasarkan koordinat *gazeX* dan *gazeY* menggunakan 8 skala *temporal window*, kemudian diekstrak fitur-fitur temporal mencakup *speed, direction, acceleration, displacement, dan standard deviation*. Model *deep learning* yang dievaluasi meliputi *LSTM, BiLSTM*, dan *TCN* dengan *5-fold cross-validation*. Hasilnya menunjukkan peningkatan dramatis: *TCN* dengan *Robust scaling* mencapai akurasi 99,51%, dengan konfigurasi optimal berupa *temporal window* 121 sampel (2 detik) dan fitur *gazeX+gazeY+speed+direction*. Penelitian ini membuktikan bahwa data *time series* perilaku digital jauh lebih akurat dibanding pendekatan yang digabungkan untuk mengidentifikasi gaya belajar, dan menjadi referensi langsung bagi penelitian ini yang mengadopsi prinsip serupa mengidentifikasi gaya belajar dari perilaku digital pengguna, namun dengan modalitas data *teks prompting chatbot* sebagai *pengganti eye tracking*.

Secara keseluruhan, studi-studi yang telah diulas menunjukkan bahwa pola *prompting* pengguna sebagai representasi karakteristik gaya belajar dalam kerangka FSLSM belum pernah dieksplorasi secara langsung. Penelitian yang ada baru membahas efektivitas *chatbot* secara umum, strategi *prompting* dari sisi teknis, atau identifikasi gaya belajar dari log aktivitas LMS, namun belum mengintegrasikan ketiganya dalam satu kerangka penelitian yang utuh. Oleh karena itu, penelitian ini diarahkan untuk mengisi celah tersebut melalui analisis berbasis data nyata mengenai hubungan antara gaya belajar mahasiswa dan pola *prompting* dalam penggunaan *chatbot* AI, guna menjadi dasar pemikiran bagi pengembangan sistem pembelajaran berbasis AI yang lebih sesuai dengan kebutuhan dan karakteristik masing-masing pengguna.

Tabel 2. 1. Perbandingan Studi Sebelumnya

No.	Penulis	Fokus utama / Topik	Metode	Hasil
1	Pratiwi, dkk. (2024)	Persepsi mahasiswa Indonesia terhadap penggunaan *ChatGPT* dalam pembelajaran Bahasa Indonesia	Pendekatan survei	*ChatGPT* dipandang efektif sebagai alat bantu belajar dan penyusunan tulisan akademik; ditemukan variasi pola interaksi (pasif vs. eksploratif) yang mengindikasikan perbedaan individual dalam preferensi belajar
2	Fauzi, dkk. (2023)	Peran *ChatGPT* dalam meningkatkan produktivitas belajar mahasiswa perguruan tinggi	Kualitatif berbasis wawancara mendalam	Penggunaan *ChatGPT* yang terarah dan terstruktur meningkatkan produktivitas lebih signifikan; ditemukan beragam pola pemanfaatan yang mencerminkan pendekatan kognitif berbeda-beda
3	Oviedo Ramirez, dkk. (2025)	Hubungan antara gaya belajar dan sumber belajar digital dalam *Adaptive Learning Systems* (ALS)	*Systematic literature review*	FSLSM merupakan model gaya belajar paling banyak digunakan dalam ALS; setiap dimensi FSLSM dapat dipetakan ke jenis sumber belajar digital tertentu sebagai landasan personalisasi pembelajaran

No.	Penulis	Fokus utama / Topik	Metode	Hasil
4	Sawalha, dkk. (2024)	Pola *prompting* mahasiswa dan pengaruhnya terhadap performa *ChatGPT*	*Mixed methods* dan *thematic analysis*	Mahasiswa mengembangkan kecenderungan *prompting* individual yang berbeda; pola *prompting* yang elaboratif dan iteratif menghasilkan respons *ChatGPT* yang lebih relevan dan mendalam
5	Güner dan Er (2025)	Interaksi mahasiswa dengan *ChatGPT* dalam pembelajaran pemrograman	Kuasieksperimental dan analisis profil	Teridentifikasi 5 profil interaksi: *AI-Reliant Code Generators, AI-Reliant Code Generator & Refiners, AI-Collaborative Coders, AI-Assisted Code Refiners,* dan *AI-Independent Coders*, yang mencerminkan perbedaan kemandirian dan strategi kognitif
6	Pritalia, dkk. (2020)	Klasifikasi gaya belajar *Sequential–Global* pada pembelajaran multimedia menggunakan *eye tracking*	*Eye tracking (GP3) + machine learning (SVM, KNN, Logistic Regression, Naive Bayes)*	Gaya belajar *Sequential–Global* berhasil diklasifikasi dari data *eye tracking* dengan akurasi tertinggi 71% (SVM); membuktikan perilaku visual dapat menjadi sinyal gaya belajar tanpa kuesioner

No.	Penulis	Fokus utama / Topik	Metode	Hasil
		dan *machine learning*		
7	Nurlatifa, dkk. (2026)	Klasifikasi gaya belajar *Sequential–Global* menggunakan data *time series eye tracking* dan *deep learning (TCN, BiLSTM, LSTM)* berdasarkan fitur temporal *gazeX, gazeY, speed, dan direction*	*Eye tracking (GP3, 60 Hz) + deep learning (TCN/BiLSTM/LSTM) dengan 5-fold cross-validation*	*TCN* dengan *Robust scaling* mencapai akurasi 99,51%; *temporal window* 2 detik (121 sampel) dan fitur *gazeX+gazeY+speed+direction* menghasilkan performa klasifikasi optimal; membuktikan data *time series* perilaku digital jauh lebih akurat dibanding pendekatan agregat

Dari seluruh penelitian yang telah dibahas, terlihat ada celah yang belum diisi. Penelitian yang membahas *ChatGPT* dalam pendidikan seperti Pratiwi dkk., Fauzi dkk., Sawalha dkk., dan Güner & Er memang menemukan adanya perbedaan pola interaksi antar mahasiswa, namun tidak menggunakan kerangka FSLSM untuk mengklasifikasikannya secara formal. Di sisi lain, penelitian yang mengklasifikasikan gaya belajar secara otomatis seperti Pritalia dkk. dan Nurlatifa dkk. menggunakan data rekaman gerakan mata, bukan data percakapan *chatbot*. Sementara Oviedo Ramirez dkk. yang membahas FSLSM dalam sistem pembelajaran digital tidak sampai pada tahap klasifikasi otomatis berbasis data perilaku. Kesenjangan ini semakin jelas apabila dilihat dari tiga dimensi

sekaligus: pertama, belum ada penelitian yang menggunakan data riwayat percakapan *chatbot* sebagai sumber data perilaku belajar; kedua, belum ada yang mengintegrasikan analisis linguistik berbasis *Natural Language Processing* (NLP) dengan kerangka FSLSM secara bersamaan; dan ketiga, belum ada yang membangun model klasifikasi otomatis gaya belajar berbasis *machine learning* yang menggunakan pola *prompting* pengguna di *ChatGPT* sebagai fitur utamanya. Dengan kata lain, belum ada penelitian yang secara langsung menghubungkan pola *prompting* pengguna di *ChatGPT* dengan klasifikasi gaya belajar berdasarkan FSLSM menggunakan *machine learning.* Penelitian ini mencoba mengisi celah tersebut dengan menjadikan riwayat percakapan *ChatGPT* sebagai sumber data perilaku digital dan FSLSM sebagai kerangka klasifikasi gaya belajarnya.

2.2. Dasar Teori

2.2.1. Machine Learning

Machine Learning (ML) adalah cabang dari kecerdasan buatan yang memungkinkan sistem komputer untuk belajar dari data dan meningkatkan performanya secara otomatis tanpa diprogram secara eksplisit untuk setiap tugas [17]. Berbeda dengan pemrograman pada umumnya yang mengandalkan aturan yang ditulis secara manual, *Machine Learning* membangun model matematis berdasarkan pola yang ditemukan dalam data pelatihan, sehingga model tersebut dapat membuat prediksi atau keputusan terhadap data baru yang belum pernah dilihat sebelumnya.

Terdapat tiga cara utama mengajarkan komputer untuk belajar dalam *Machine Learning*. Pertama, *supervised learning*, yaitu model dilatih menggunakan data berlabel di mana setiap input memiliki *output* yang sudah diketahui. Kedua, *unsupervised learning*, yaitu model mencari pola atau struktur tersembunyi dalam data yang tidak memiliki label. Ketiga, *semi-supervised learning*, yaitu kombinasi dari keduanya di mana hanya sebagian data yang berlabel. Penelitian ini menggunakan pendekatan *supervised learning* karena setiap data pertanyaan mahasiswa sudah dipasangkan dengan jawaban yang jelas, yaitu label gaya belajar yang didapat dari kuesioner ILS, sehingga komputer

bisa belajar dari contoh-contoh yang sudah diketahui kebenarannya (*ground truth)*. Tugas yang dilakukan adalah klasifikasi, yaitu memprediksi kategori gaya belajar (misalnya *Active* atau *Reflective*) berdasarkan fitur-fitur yang diekstrak dari teks *prompt*.

Dalam klasifikasi teks, *Machine Learning* bekerja melalui dua tahap utama. Tahap pertama adalah ekstraksi fitur, yaitu proses mengubah teks mentah menjadi deretan angka-angka yang bisa dibaca dan diolah oleh komputer. Tahap kedua adalah pelatihan model, yaitu algoritma mempelajari pola hubungan antara angka-angka hasil konversi teks tadi dengan kategori yang ingin diprediksi [17]. Pemilihan algoritma klasifikasi yang tepat sangat memengaruhi performa model, sehingga dalam penelitian ini dilakukan perbandingan tiga algoritma untuk menentukan algoritma yang paling efektif, yaitu *Random Forest, Support Vector Machine,* dan *Naive Bayes*.

2.2.2. *Natural Language Processing* (NLP)

Natural Language Processing (NLP) adalah cabang kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa manusia. NLP memungkinkan komputer untuk memahami, menginterpretasikan, dan menghasilkan teks dalam bahasa alami secara otomatis. Dalam konteks penelitian ini, NLP digunakan untuk memproses data teks *prompt* mahasiswa menjadi angka angka yang dapat digunakan sebagai *input* model *machine learning* [18].

Pipeline NLP yang diterapkan dalam penelitian ini terdiri dari beberapa tahapan berurutan. Pertama, *text cleaning* yaitu proses pembersihan teks dari elemen yang tidak relevan seperti *URL*, karakter khusus, tanda baca berlebih, dan normalisasi spasi. Kedua, *lowercasing* yaitu mengubah seluruh teks ke huruf kecil untuk memastikan konsistensi. Ketiga, *tokenisasi* yaitu pemecahan teks menjadi unit-unit kata atau token. Keempat, *stopword removal* yaitu penghapusan kata-kata yang tidak membawa makna signifikan seperti "yang", "dan", "di", "*the*", "*is*", dan sebagainya. Kelima, normalisasi token numerik yaitu penggantian angka dengan token khusus NUM untuk mengurangi dimensi fitur[19].

2.2.3. TF-IDF (*Term Frequency-Inverse Document Frequency*)

TF-IDF adalah metode yang digunakan untuk mengukur seberapa penting sebuah kata dalam suatu teks, lalu mengubah teks tersebut menjadi deretan angka yang bisa diproses komputer. Metode ini mengukur seberapa penting suatu kata dalam sebuah dokumen dibandingkan kumpulan dokumen yang ada [17]. Secara konseptual, TF-IDF menggabungkan dua pertimbangan sekaligus. Pertama, seberapa sering sebuah kata muncul dalam satu dokumen, kata yang lebih sering muncul dianggap lebih relevan terhadap dokumen tersebut. Kedua, seberapa jarang kata tersebut muncul di seluruh dokumen dalam kumpulan data yang digunakan.

Kata yang muncul di hampir semua dokumen dianggap kurang informatif dan diberi bobot lebih rendah. Kombinasi keduanya menghasilkan bobot yang tinggi untuk kata-kata yang sering muncul dalam satu dokumen tertentu namun jarang ditemukan di dokumen lain, sehingga kata tersebut dianggap sebagai ciri khas dokumen tersebut. Dalam penelitian ini, TF-IDF digunakan untuk menangkap makna dan isi dari keseluruhan teks *prompt* setiap partisipan. Detail konfigurasi dan penerapannya dijelaskan lebih lanjut pada Bab 3.

2.2.4. Algoritma Klasifikasi

Algoritma klasifikasi adalah metode dalam *supervised learning* yang bertugas mengelompokkan data masukan ke dalam salah satu kategori yang sudah ditentukan sebelumnya [17]. Dalam penelitian ini, tiga algoritma klasifikasi dibandingkan untuk menentukan pendekatan yang paling efektif dalam mengklasifikasikan gaya belajar mahasiswa berdasarkan pola *prompting*.

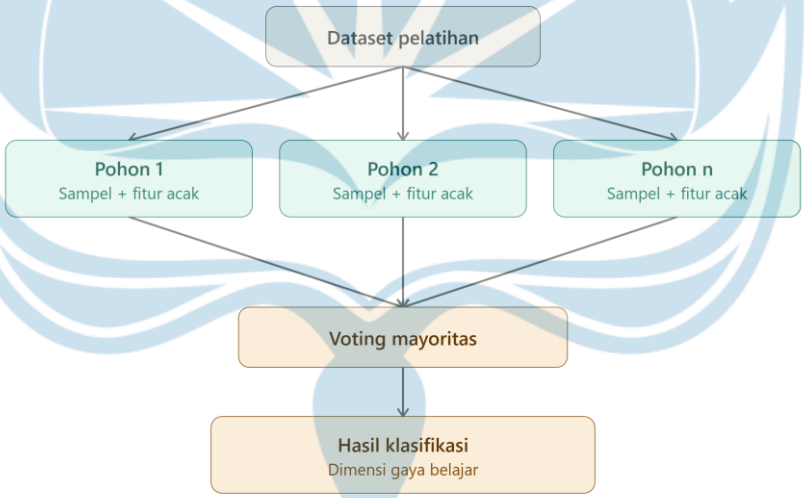
2.2.4.1. Random Forest

Random Forest adalah algoritma *ensemble learning* yang bekerja dengan membangun sejumlah pohon keputusan (*decision tree*) secara paralel pada subset data yang

berbeda, kemudian menggabungkan prediksi seluruh pohon melalui mekanisme *majority voting[20]*. Algoritma ini diperkenalkan oleh Breiman dan terbukti efektif untuk data berdimensi tinggi seperti fitur teks.

Proses pembangunan *Random Forest* melibatkan dua sumber keacakan. Pertama, *bootstrap aggregating* (*bagging*), yaitu setiap pohon dilatih pada subset data yang dipilih secara acak dengan pengembalian dari dataset asli. Kedua, *random feature selection*, yaitu pada setiap *node* pemisahan, hanya sebagian fitur yang dipilih secara acak sebagai kandidat pemisahan terbaik. Kedua mekanisme ini menghasilkan pohon-pohon yang beragam satu sama lain sehingga mengurangi risiko *overfitting*.

Kelebihan *Random Forest* yang relevan dengan penelitian ini antara lain mampu menangani data dengan jumlah fitur yang sangat besar (>300 fitur gabungan linguistik dan TF-IDF), tahan terhadap *overfitting* dibandingkan *decision tree* tunggal, serta mampu mengestimasi pentingnya setiap fitur (*feature importance*) sehingga dapat diidentifikasi fitur *prompting* yang paling berkontribusi terhadap klasifikasi gaya belajar.



Gambar 2. 1. Ilustrasi Cara Kerja Algoritma *Random Forest*

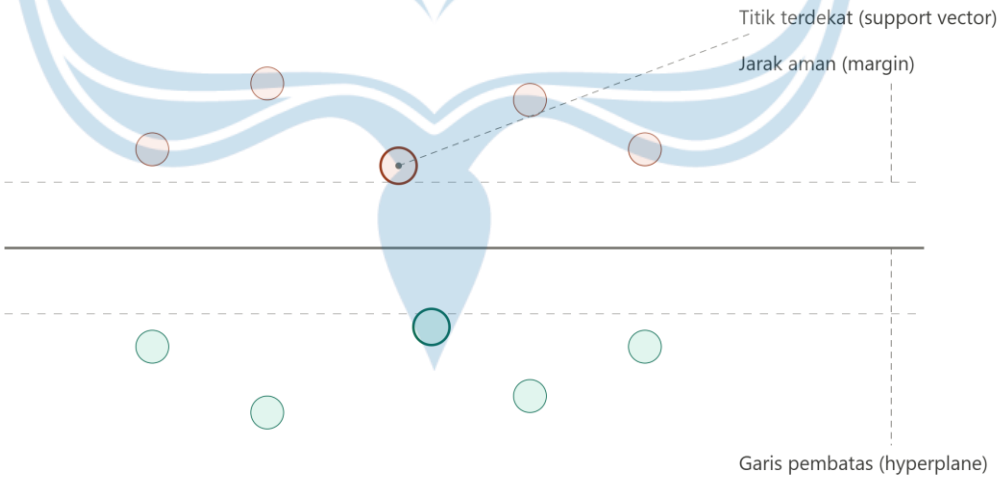
Random Forest seperti bertanya ke banyak orang sebelum mengambil keputusan. Setiap orang (pohon) hanya melihat sebagian informasi dan memberi jawabannya sendiri, lalu jawaban yang paling banyak dipilih menjadi keputusan akhir. Cara ini membuat hasilnya lebih bisa dipercaya dibanding hanya bertanya ke satu orang saja.

2.2.4.2. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma klasifikasi yang bekerja dengan mencari *hyperplane* optimal yang memisahkan dua kelas dengan margin terbesar dalam ruang fitur berdimensi tinggi . Konsep utama *Support Vector Machine* adalah memaksimalkan jarak (*margin*) antara *hyperplane* pemisah dan titik data terdekat dari masing-masing kelas, yang disebut *support vectors*[21].

Untuk data yang tidak dapat dipisahkan secara linear, *Support Vector Machine* menggunakan teknik *kernel trick* yang memetakan data ke ruang berdimensi lebih tinggi sehingga pemisahan linear menjadi memungkinkan. Kernel yang umum digunakan antara lain kernel linear, RBF (*Radial Basis Function*), dan polinomial. Dalam penelitian ini, *kernel RBF (Radial Basis Function)* digunakan karena memberikan performa lebih baik dibandingkan *kernel linear* setelah proses *GridSearchCV*.

Kelebihan *Support Vector Machine* dalam penelitian ini adalah kemampuannya bekerja dengan baik pada dataset berukuran kecil hingga menengah dan ruang fitur berdimensi tinggi, yang sesuai dengan karakteristik dataset penelitian ini yang terdiri dari 96 partisipan dengan lebih dari 300 fitur gabungan.



Gambar 2. 2. Ilustrasi Cara Kerja Algoritma *Support Vector Machine*(SVM)

Support Vector Machine bekerja dengan mencari garis pembatas yang paling adil antara dua kelompok data. Garis ini ditarik sejauh mungkin dari titik-titik yang posisinya paling

dekat dengan kelompok lawan, supaya jarak amannya paling besar dan pemisahan tidak mudah salah ketika ada data baru.

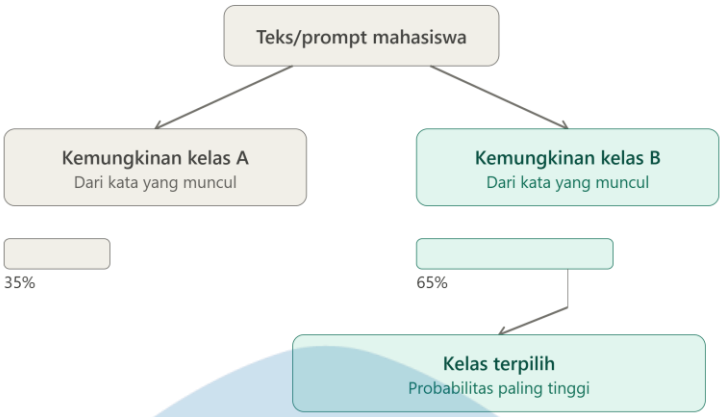
2.2.4.3. Naive Bayes

Naive Bayes adalah algoritma klasifikasi probabilistik yang didasarkan pada *Teorema Bayes* dengan asumsi kuat (*naive*) bahwa setiap fitur bersifat independen satu sama lain *given label* kelasnya . Meskipun asumsi independensi ini jarang terpenuhi secara sempurna dalam data nyata, *Naive Bayes* tetap menunjukkan performa yang kompetitif dalam banyak tugas klasifikasi teks[22].

Untuk data teks, varian yang umum digunakan adalah *Multinomial Naive Bayes*, yang menghitung probabilitas kemunculan setiap kata dalam suatu kelas berdasarkan frekuensinya. Model ini kemudian memprediksi kelas dengan probabilitas tertinggi berdasarkan kombinasi probabilitas fitur-fitur yang ada.

Kelebihan *Naive Bayes* dalam penelitian ini adalah efisiensi komputasional yang sangat tinggi bahkan pada dataset berdimensi tinggi, serta kemampuannya bekerja dengan baik pada *dataset* kecil karena hanya membutuhkan sedikit data pelatihan untuk mengestimasi parameter yang diperlukan. Hal ini relevan mengingat jumlah partisipan dalam penelitian ini terbatas.

Varian yang digunakan dalam penelitian ini adalah *Complement Naive Bayes* (CNB), yang merupakan pengembangan dari *Multinomial Naive Bayes* dan terbukti lebih baik pada data dengan *class imbalance* karena menghitung probabilitas dari komplemen kelas, sehingga estimasi parameter kelas minoritas lebih akurat.



Gambar 2. 3. Ilustrasi Cara Kerja Algoritma *Naive Bayes*

Naive Bayes bekerja seperti menebak berdasarkan kata-kata yang muncul. Setiap kata dalam teks memberi sedikit "petunjuk" yang menambah kemungkinan teks itu termasuk kelas tertentu. Setelah semua petunjuk dihitung, kelas dengan total kemungkinan paling besar dipilih sebagai jawaban akhir.

2.2.4.4. Perbandingan dan Alasan Pemilihan

Ketiga algoritma di atas dipilih sebagai pembanding karena masing-masing merepresentasikan pendekatan yang berbeda secara fundamental: *Random Forest* berbasis *ensemble* pohon keputusan, *Support Vector Machine* berbasis geometri *hyperplane*, dan *Naive Bayes* berbasis probabilistik. Perbandingan ketiganya memberikan gambaran tentang pendekatan mana yang paling sesuai untuk karakteristik data penelitian ini.

Berdasarkan studi terdahulu, Pritalia, dkk. [15] menggunakan *Support Vector Machine* dan KNN untuk klasifikasi gaya belajar berbasis *eye tracking* dan memperoleh akurasi tertinggi dengan *Support Vector Machine* sebesar 71%. Nurlatifa, dkk. [16] kemudian menunjukkan bahwa pendekatan berbasis *ensemble* dan *deep learning* mampu melampaui performa *Support Vector Machine* secara signifikan pada data perilaku digital yang serupa. Oleh karena itu, *Random Forest* dipilih sebagai algoritma utama dalam penelitian ini karena kemampuannya menangani fitur berdimensi tinggi, menghasilkan *feature importance* yang interpretatif, dan terbukti lebih stabil dibandingkan *decision tree* tunggal pada dataset berukuran kecil hingga menengah.

2.2.5. SMOTE (*Synthetic Minority Over-sampling Technique*)

SMOTE (*Synthetic Minority Over-sampling Technique*) adalah teknik *oversampling* yang dikembangkan oleh Chawla, dkk. [23] untuk menangani masalah ketidakseimbangan kelas (*class imbalance*) dalam dataset klasifikasi. Berbeda dengan *oversampling* sederhana yang hanya menduplikasi data minoritas, *SMOTE* menghasilkan sampel sintetis baru dengan cara interpolasi antara data minoritas yang sudah ada. Cara kerja *SMOTE* adalah sebagai berikut: untuk setiap sampel minoritas, *SMOTE* mencari k-tetangga terdekat (*k-nearest neighbors*) di kelas yang sama, lalu membuat sampel baru di sepanjang garis yang menghubungkan sampel tersebut dengan salah satu tetangganya. Proses ini dilakukan secara acak pada *feature space* sehingga menghasilkan data sintetis yang bervariasi, bukan sekadar salinan. Dalam penelitian ini, *SMOTE* diterapkan pada *data training* sebelum pelatihan model untuk menyeimbangkan distribusi kelas pada setiap dimensi ILS, mengingat distribusi label gaya belajar tidak seimbang secara alami. Penerapan *SMOTE* dilakukan hanya pada data training (bukan data uji) untuk mencegah data leakage yang dapat mengakibatkan evaluasi yang terlalu optimis.

Dalam penelitian ini, strategi *oversampling* dipilih secara adaptif berdasarkan tingkat ketidakseimbangan kelas pada setiap dimensi. Untuk dimensi *dengan rasio imbalance* di bawah 5x, *SMOTE* digunakan sebagaimana dijelaskan di atas.

Selain itu, untuk mengurangi risiko *noise* dan tumpang tindih (*overlap*) antar kelas yang dapat timbul akibat proses *oversampling*, penelitian ini menggunakan kombinasi *SMOTE* dengan *Tomek Links*, yang dikenal sebagai *SMOTE-Tomek*. *Tomek Links* bekerja dengan mengidentifikasi pasangan sampel dari kelas yang berbeda yang saling menjadi tetangga terdekat satu sama lain (*nearest neighbor pair*), kemudian menghapus sampel kelas majoritas dari pasangan tersebut. Dengan demikian, *SMOTE-Tomek* tidak hanya menambah jumlah sampel minoritas, tetapi juga membersihkan batas antar kelas sehingga distribusi data hasil *resampling* menjadi lebih bersih dan tidak rentan terhadap sampel yang ambigu.

2.2.6. Korelasi *Spearman*

Korelasi *Spearman* adalah metode statistik non-parametrik yang digunakan untuk mengukur kekuatan dan arah hubungan antara dua variabel. Berbeda dengan korelasi *Pearson* yang mengasumsikan data berdistribusi normal, korelasi *Spearman* bekerja berdasarkan peringkat data sehingga lebih *robust* terhadap data yang tidak berdistribusi normal maupun data yang mengandung *outlier*.

Nilai koefisien korelasi *Spearman* berkisar antara -1 hingga +1. Nilai mendekati +1 menunjukkan hubungan positif yang kuat, nilai mendekati -1 menunjukkan hubungan negatif yang kuat, dan nilai mendekati 0 menunjukkan tidak adanya hubungan yang signifikan antara kedua variabel. Dalam penelitian ini, korelasi *Spearman* digunakan untuk menganalisis kekuatan hubungan antara fitur-fitur linguistik *prompting* dengan skor gaya belajar pada setiap dimensi FSLSM.

2.2.7. Gaya Belajar dan Model *Felder-Silverman* (FSLSM)

Gaya belajar merujuk pada preferensi individu dalam menerima, mengolah, dan memahami informasi. Oviedo Ramirez, dkk. (2025) mendefinisikan gaya belajar sebagai pola perilaku dan pemikiran yang khas yang memengaruhi cara individu memproses informasi. Terdapat sejumlah model gaya belajar yang telah dikembangkan, di antaranya model *Kolb* (1984), *VARK* (Fleming, 1987), *Honey* dan *Mumford* (1986), serta model *Felder-Silverman* (1988). Di antara model-model tersebut, *Felder-Silverman Learning Styles Model* (FSLSM) [24]adalah yang paling banyak digunakan dalam pengembangan *Adaptive Learning Systems* (ALS).

FSLSM mengklasifikasikan gaya belajar ke dalam empat dimensi dengan dua kutub pada setiap dimensinya, sebagaimana disajikan pada tabel berikut:

Tabel 2. 2. Dimensi dan Tipologi dalam FSLSM

Dimensi	Tipologi	Karakteristik Utama
Persepsi (*Perception*)	*Sensory* ↔ *Intuitive*	*Sensory*: menyukai fakta konkret, data spesifik, prosedur. *Intuitive*: menyukai konsep abstrak, pola, dan teori.
Representasi (*Input*)	*Visual* ↔ *Verbal*	*Visual*: belajar melalui gambar, diagram, flowchart. *Verbal*: belajar melalui penjelasan tertulis/lisan.
Pemrosesan (*Processing*)	*Active* ↔ *Reflective*	*Active*: belajar dengan melakukan, diskusi, eksperimen. *Reflective*: belajar dengan merenung, menganalisis secara mendalam.
Pemahaman (*Understanding*)	*Sequential* ↔ *Global*	*Sequential*: memproses informasi secara bertahap dan linear. *Global*: menangkap gambaran besar sebelum detail.

Instrumen pengukuran FSLSM adalah *Index of Learning Styles* (ILS), yaitu kuesioner 44 item yang menghasilkan nilai preferensi pada setiap dimensi[24] Ayyoub dan Al-Kadi [13] dalam penelitiannya memvalidasi pendekatan identifikasi gaya belajar menggunakan data perilaku digital nyata dan menemukan bahwa FSLSM memiliki kekuatan prediktif tertinggi dibandingkan model gaya belajar lainnya dalam konteks sistem *e-learning* adaptif, khususnya karena kemampuannya memetakan perilaku pengguna secara kuantitatif dan multidimensional.

2.2.8. Konsep *Prompting* dalam Interaksi *Chatbot AI*

Prompting merupakan proses pemberian instruksi, pertanyaan, atau perintah kepada sistem AI untuk menghasilkan respons tertentu. Dalam konteks pembelajaran, *prompting* mencerminkan strategi pengguna dalam menggali informasi dan membangun pemahaman. Sawalha, dkk. dalam studinya terhadap mahasiswa dengan ratusan percakapan *ChatGPT* menemukan bahwa mahasiswa mengembangkan kecenderungan *prompting* individual yang berbeda-beda, sebagian melakukan elaborasi bertahap, sebagian meminta jawaban langsung, dan sebagian melakukan uji coba penyempurnaan. Pola-pola ini bukan sekadar perbedaan teknis, melainkan mencerminkan pendekatan kognitif yang berbeda dalam memecahkan masalah[7].

H. Güner and E. Er [8]memperkuat temuan ini dengan mengidentifikasi lima profil interaksi mahasiswa dengan *ChatGPT*, yaitu: *AI-Reliant Code Generators*, *AI-Reliant Code Generator & Refiners*, *AI-Collaborative Coders*, *AI-Assisted Code Refiners*, dan *AI-Independent Coders*. Profil-profil ini menggambarkan spektrum kemandirian dan strategi kolaborasi yang beragam, yang secara implisit mencerminkan perbedaan preferensi belajar. Xu, dkk. juga mengindikasikan bahwa intensitas dan pola penggunaan *chatbot* berpengaruh terhadap perilaku belajar mahasiswa[25], sehingga cara mahasiswa melakukan *prompting* dapat menjadi indikator strategi belajar mereka. Dengan demikian, *prompting* dapat diposisikan sebagai representasi perilaku belajar dalam konteks interaksi manusia–AI[26].

2.2.9. Gaya Belajar dalam Konteks Pembelajaran Digital dan *Chatbot AI*

Dalam pembelajaran digital, gaya belajar tidak lagi hanya terlihat dari cara membaca atau mendengarkan materi, tetapi juga dari cara individu berinteraksi dengan teknologi. Oviedo Ramirez, dkk. [11] menegaskan bahwa ketika sumber belajar digital diselaraskan

dengan preferensi gaya belajar individu, mahasiswa lebih mungkin mencapai pemahaman yang lebih dalam dan retensi informasi yang lebih kuat. Profil tiap dimensi FSLSM[24]yang relevan dengan pola *prompting* adalah sebagai berikut.

Mahasiswa dengan gaya *Active* diasumsikan belajar melalui keterlibatan langsung, diskusi, dan eksperimentasi. Dalam konteks *prompting*, mereka diasumsikan membuat prompt yang mengajak chatbot berdiskusi atau meminta contoh penerapan secara eksploratif. Sebaliknya, mahasiswa dengan gaya *Reflective* diasumsikan memproses informasi secara internal dan mendalam, pola *prompting* mereka diperkirakan lebih analitis dan bertahap, mengajukan pertanyaan lanjutan sebelum beralih topik. Mahasiswa *Sequential* memproses informasi secara linear sehingga pola *prompting* mereka diasumsikan mengikuti urutan yang sistematis, sedangkan mahasiswa *Global* diasumsikan memulai dengan pertanyaan yang luas lalu mempersempit fokus[24]. Mahasiswa *Sensing* diasumsikan membuat *prompt* yang spesifik dan berbasis masalah praktis, sementara mahasiswa *Intuitive* diasumsikan lebih eksploratoris dan terbuka. Preferensi *Visual* vs. *Verbal* diasumsikan dapat tercermin dalam jenis konten yang diminta, apakah mahasiswa meminta tabel dan poin-poin terstruktur, atau penjelasan teks panjang.

Ayyoub dan Al-Kadi [13] menunjukkan bahwa identifikasi gaya belajar dapat dilakukan secara otomatis melalui analisis pola aktivitas digital mahasiswa menggunakan pendekatan *semi-supervised machine learning*. Pritalia [15]memperkuat pendekatan ini dengan menggunakan *eye-tracking* dan *machine learning* untuk mengklasifikasi gaya belajar global dan sekuensial pada *multimedia learning*. Nurlatifa, dkk. [16] melanjutkan dengan membuktikan bahwa gaya belajar *sequential-global* dapat diklasifikasikan secara akurat melalui data *eye-tracking* menggunakan *deep learning*. Temuan-temuan ini menunjukkan bahwa aktivitas digital secara umum punya potensi jadi sinyal tersembunyi untuk mengenali gaya belajar seseorang. Pola *prompting* juga termasuk salah satu bentuk aktivitas digital, jadi ada kemungkinan pola *prompting* bisa berfungsi serupa. Hanya saja, sampai saat ini belum ada penelitian yang secara khusus menguji hal ini dalam konteks *prompting ke chatbot AI*.

Beberapa asumsi yang sudah dijelaskan di paragraf sebelumnya sebagian sejalan dengan hasil analisis korelasi Spearman pada penelitian ini. Pada dimensi *Sequential/Global*, fitur *kw_score_sequential* berkorelasi negatif dengan label *Sequential* (ρ = −0,3719), yang cocok dengan asumsi bahwa mahasiswa *Global* lebih sering minta bantuan menyusun struktur dari *chatbot*. Pada dimensi *Sensing/Intuitive*, fitur *short_prompt_ratio* berkorelasi positif ke arah *Sensing* (ρ = +0,345), yang sejalan dengan asumsi bahwa mahasiswa *Sensing* cenderung membuat *prompt* yang singkat dan langsung ke inti masalah. Sementara itu, pada dimensi *Visual/Verbal*, sinyal pendukungnya memang ada, tapi tidak selalu muncul dari fitur yang sesuai dugaan, melainkan dari cara mahasiswa bertanya secara umum. Pada dimensi *Active/Reflective*, dukungan hasil pengujiannya merupakan yang paling lemah dibanding dimensi lain, hanya dua fitur yang signifikan secara statistik, sehingga asumsi pada dimensi ini masih perlu dipandang sebagai yang paling belum terbukti. Secara keseluruhan, kekuatan korelasi yang ditemukan masih tergolong lemah sampai sedang, sehingga keterkaitan antara gaya belajar dan pola *prompting* ini sebaiknya dilihat sebagai langkah awal (*proof of concept*) yang masih perlu diuji lebih lanjut, bukan sebagai kesimpulan yang sudah pasti benar.

2.2.10. *Adaptive Learning System* (ALS) dan Peran Gaya Belajar

Adaptive Learning System (ALS) adalah sistem pembelajaran yang mampu memantau aktivitas pengguna, menginterpretasikan data interaksi, menginferensikan kebutuhan dan preferensi pengguna, dan bertindak sesuai pengetahuan yang diperoleh untuk memfasilitasi proses pembelajaran secara dinamis. Oviedo Ramirez, dkk. [11]menjelaskan bahwa ALS dapat memanfaatkan FSLSM sebagai dasar rekomendasi konten digital, menyesuaikan jenis sumber belajar berdasarkan profil gaya belajar pengguna. Ayyoub dan Al-Kadi [13] memperkuat argumentasi ini dengan menunjukkan bahwa dimensi FSLSM dapat dipetakan secara langsung ke indikator perilaku yang dapat dilacak oleh sistem digital.

Dalam konteks penelitian ini, *chatbot* AI dapat dipandang sebagai komponen interaktif dari ekosistem pembelajaran digital. Pola *prompting* mahasiswa dalam berinteraksi dengan *chatbot* berpotensi menjadi data perilaku yang mencerminkan preferensi gaya belajar, sehingga dapat menjadi *proof of concept* awal sebelum dieksplorasi lebih lanjut sebagai komponen deteksi profil dalam pengembangan ALS yang lebih personal dan responsif terhadap kebutuhan individu.

2.2.11. Kerangka Konseptual Penelitian

Berdasarkan uraian teori di atas, kerangka konseptual penelitian ini memposisikan:

- Gaya belajar (FSLSM) sebagai variabel karakteristik individu
- Pola *prompting* sebagai representasi strategi belajar dalam interaksi AI
- Potensi deteksi gaya belajar dari pola *prompting* sebagai *proof of concept* awal yang dapat menjadi pijakan eksplorasi lebih lanjut ke arah sistem pembelajaran adaptif

Dengan memahami hubungan antara gaya belajar dan pola *prompting*, penelitian ini diharapkan dapat memberikan landasan konseptual awal yang membuka kemungkinan perancangan sistem pembelajaran berbasis AI yang lebih personal dan selaras dengan karakteristik individu mahasiswa, meskipun penerapan teknisnya masih memerlukan penelitian lanjutan. Secara keseluruhan, tinjauan pustaka ini menunjukkan bahwa meskipun penelitian tentang *chatbot* AI dalam pendidikan telah berkembang pesat, integrasi antara analisis pola *prompting* dan identifikasi gaya belajar berbasis FSLSM masih merupakan area yang belum banyak dieksplorasi. Penelitian ini hadir untuk mengisi kesenjangan tersebut dengan menganalisis keterkaitan antara kedua variabel tersebut dalam konteks mahasiswa perguruan tinggi di Indonesia.