

BAB II

TINJAUAN PUSTAKA

Sudah banyak penelitian terkait dengan klasifikasi dengan menggunakan algoritma Random Forest, maupun sistem rekomendasi makanan dengan menggunakan beberapa algoritma *machine learning*, salah satunya adalah K-Nearest Neighbors. Dari penelitian-penelitian sebelumnya, dapat digunakan sebagai pedoman dalam pengembangan model yang akan dibuat. Sistem rekomendasi berkembang pesat di berbagai bidang, dengan banyak yang memanfaatkan *machine learning* untuk memberikan rekomendasi yang lebih akurat. Dalam konteks makanan, sistem rekomendasi membantu pengguna menemukan pilihan makanan yang sesuai dengan kebutuhan dan preferensi pribadi pengguna.

Penelitian pertama dengan judul “Movie Recommendation System Using Cosine Similarity and KNN” bertujuan untuk membantu pengguna dalam menemukan film yang sesuai dengan preferensi mereka. Penelitian ini didasari oleh perkembangan internet yang memicu meningkatnya tingkat produksi film yang membuat pengguna kesulitan untuk menemukan film yang sesuai dengan preferensi mereka. Sistem rekomendasi berbasis *machine learning* menjadi solusi dari permasalahan tersebut dengan menggunakan algoritma KNN yang didukung oleh *Cosine Similarity* untuk menemukan film-film yang serupa dengan yang ditonton oleh pengguna sebelumnya. Sistem rekomendasi tersebut telah berhasil dikembangkan dan menunjukkan hasil yang memuaskan dalam memberikan rekomendasi film yang sesuai dengan preferensi pengguna [10].

Penelitian kedua dengan judul “The Research for Recommendation System Based on Improved KNN Algorithm” bertujuan untuk meneliti sistem rekomendasi secara umum dengan menggunakan algoritma KNN. Dilatarbelakangi oleh berkembangnya teknologi dan kesulitan pengguna dalam menemukan informasi yang relevan di tengah banyaknya data yang ada, sistem rekomendasi menjadi solusi yang tepat untuk masalah tersebut. Penelitian

fokus kepada algoritma KNN dan peningkatan algoritma KNN dengan cara melakukan kompresi data yakni mengurangi data yang jarang muncul untuk meningkatkan akurasi prediksi, dan efek global yang mempertimbangkan karakteristik dari pengguna dan objek. Eksperimen dilakukan menggunakan kumpulan data dari MovieLens dengan tiga skala ukuran yakni 100K, 1M, dan 10M. Hasil eksperimen dievaluasi menggunakan *Root Mean Squared Error* (RMSE) yang menunjukkan bahwa algoritma KNN yang ditingkatkan menunjukkan kinerja yang lebih baik dibandingkan KNN standar, lalu tingkat presisi rekomendasi naik sebesar 2,3% pada Kumpulan data dengan jumlah 1M [11].

Penelitian ketiga dengan judul “K-Nearest Neighbors Method for Recommendation System in Bangkalan’s Tourism” bertujuan untuk membuat sistem rekomendasi untuk destinasi wisata di pulau Bangkalan. Menjadi solusi dari permasalahan banyaknya objek wisata yang ada pada pulau Bangkalan yang membuat wisatawan bingung dalam memilih objek wisata yang sesuai. Penelitian kali ini menggunakan algoritma KNN untuk memberikan rekomendasi berdasarkan karakteristik wisatawan, data Dinas Pariwisata Bangkalan yang terdiri dari 315 rekaman data digunakan untuk melatih model *machine learning*. Sebelum diterapkan algoritma KNN, dilakukan *preprocessing* terlebih dahulu pada data lalu lanjut dengan pengimplementasian algoritma KNN yang diawali dengan menghitung jarak menggunakan *Euclidean Distance* lalu menentukan jumlah tetangga terdekat (k) untuk dijadikan rekomendasi objek wisata. Model dievaluasi dengan menggunakan akurasi rekomendasi sebagai metrik utama yang diperoleh tingkat akurasi sebesar 78%, menunjukkan bahwa model sudah cukup baik tetapi masih harus dikembangkan seperti dengan cara mempertimbangkan faktor eksternal seperti tren media sosial dan ulasan pengguna [12].

Penelitian keempat dengan berjudul “Penerapan Algoritma k-Nearest Neighbors pada Aplikasi Rekomendasi Makanan Sehat Berbasis Mobile” bertujuan untuk membahas penerapan algoritma KNN dalam aplikasi rekomendasi makanan sehat berbasis mobile. Penelitian ini mengangkat isu

kesulitan masyarakat dalam memilih makanan sehat yang sesuai disesuaikan dengan alergi, kondisi kesehatan, dan kebutuhan nutrisi. Algoritma yang digunakan adalah KNN dan metrik *cosine distance*, dengan menjadikan *website* FatSecret sebagai data makanan untuk model *machine learning* ini. Data tersebut melewati *preprocessing* data berdasarkan waktu makan, kondisi kesehatan, dan alergi makanan, lalu dilakukan penerapan algoritma KNN menggunakan *cosine distance* untuk menentukan kesamaan antara kebutuhan nutrisi pengguna dengan data makanan yang tersedia. Sistem berhasil untuk memberikan rekomendasi makanan dengan hasil pengujian yang menunjukkan bahwa makanan dengan *cosine distance* mendekati 0 memiliki kesesuaian nutrisi yang tinggi dengan pengguna, maka dapat disimpulkan algoritma KNN dan *cosine distance* efektif dalam membantu pengguna untuk memilih makanan yang sesuai [13].

Penelitian kelima dengan judul “Metode Machine Learning untuk Klasifikasi Data Gizi Balita dengan Algoritma Naïve Bayes, KNN dan Decision Tree” bertujuan untuk klasifikasi data gizi balita dengan menggunakan algoritma Naïve Bayes, KNN, dan *Decision Tree*. Latar belakang dari penelitian ini mengangkat isu kondisi stunting pada balita yang terjadi karena beberapa hal seperti kondisi sosial ekonomi, gizi ibu saat hamil, penyakit bayi, dan pola makan yang buruk. *Machine Learning* menjadi solusi dari permasalahan tersebut dengan mampu untuk mengklasifikasikan data gizi balita sehingga dapat dilakukan intervensi sejak dini. Data yang digunakan pada penelitian ini berasal dari posyandu, setelah data diperoleh, dilakukan *preprocessing* data yang terdiri dari data *cleaning*, transformasi data, dan data *split* menjadi 70% pelatihan dan 30% pengujian. Setelah itu, diterapkan ketiga algoritma *machine learning* tersebut dan mendapatkan hasil perbandingan untuk *Naïve Bayes* memiliki akurasi sebesar 23%, untuk KNN memiliki akurasi sebesar 40% dan untuk *Decision Tree* memiliki akurasi sebesar 48%, dari hasil perbandingan tersebut dapat ditarik kesimpulan bahwa *Decision Tree* merupakan algoritma yang terbaik tetapi tetap perlu dikembangkan sehingga model dapat bekerja dengan lebih baik [14].

Penelitian keenam dengan judul "Nudging Towards Health? Examining the Merits of Nutrition Labels and Personalization in a Recipe Recommender System" bertujuan untuk membahas bagaimana sistem rekomendasi makanan dapat membantu pengguna untuk memilih resep makanan yang lebih sehat menggunakan sistem rekomendasi seperti *content-based filtering* dan *collaborative content filtering* yang dipadukan dengan teknik *nudging* atau sebuah dorongan halus untuk mengarahkan keputusan individu tanpa memaksa atau membatasi pilihan mereka. Penelitian ini untuk menguji apakah label nutrisi (seperti Nutri-Score dan Multiple Traffic Light) dalam sistem rekomendasi resep dapat mendorong pengguna untuk memilih resep yang lebih sehat. Dan diperoleh bahwa personalisasi dalam sistem rekomendasi makanan dapat memperkuat kebiasaan memilih makanan yang tidak sehat. Sementara itu, label nutrisi membantu mempermudah pemilihan tetapi tidak selalu mengubah keputusan pengguna [15].

Penelitian ketujuh dengan judul "Random Forest Classification Method for Predicting Intertidal Wetland Migration Under Sea Level Rise" bertujuan untuk memprediksi distribusi masa depan dari lahan basah intertidal seperti hutan mangrove dan saltmarsh akibat kenaikan permukaan laut. Penelitian dilakukan dengan menggunakan algoritma Random Forest dan menggunakan dataset yang berdasarkan data hidro-geomorfologis seperti *Digital Elevation Model* (DEM) dan peta vegetasi makrofit. Model menunjukkan performa yang kurang baik walaupun menunjukkan kemampuan prediktif yang baik terhadap distribusi lahan basah saat ini di banyak estuari, terdapat beberapa kendala seperti kesalahan klasifikasi antara mangrove dan saltmarsh, serta keterbatasan resolusi data. Evaluasi dilakukan melalui perbandingan prediksi model terhadap kondisi aktual di lapangan [16].

Penelitian kedelapan dengan judul "Improved Random Forest Classification Model Combined with C5.0 Algorithm for Vegetation Feature Analysis in Non-Agricultural Environments" bertujuan untuk meningkatkan akurasi klasifikasi vegetasi pada habitat buatan non-pertanian dengan memanfaatkan algoritma Random Forest yang telah disempurnakan melalui

integrasi dengan algoritma C5.0. Dataset yang digunakan pada penelitian ini diperoleh dengan menggunakan citra satelit resolusi tinggi dari GF-2. Hasil penelitian menunjukkan bahwa model Random Forest yang ditingkatkan berhasil mencapai akurasi keseluruhan sebesar 87% dan koefisien Kappa sebesar 74%, jauh lebih baik dibandingkan model awal yang memiliki akurasi keseluruhan 64% [17].

Penelitian kesembilan dengan judul “Healthcare Recommender System Using Random Forest Classifier” bertujuan untuk mengembangkan sebuah sistem rekomendasi obat yang dispesifikan untuk mempertimbangkan kondisi setiap pengguna. Penelitian ini menggunakan algoritma Random Forest, dataset yang digunakan memuat data-data seperti informasi gejala, penyakit, dan obat yang sesuai. Dataset diproses melalui melalui tahap pembersihan data dan normalisasi agar dapat dilanjutkan ke tahap pelatihan model dengan lebih optimal. Sistem ini memungkinkan pengguna untuk memasukkan gejala yang mereka alami, dan berdasarkan data tersebut, model akan mengidentifikasi dan mengeluarkan output berupa obat yang sesuai. Hasil pengujian menunjukkan tingkat akurasi sebesar 90% membuktikan Random Forest cukup efektif dalam melakukan klasifikasi [18].

Penelitian kesepuluh dengan judul “Application of the Random Forest Algorithm for Predicting Hajj Registration Numbers at Kemenag Lhokseumawe” bertujuan untuk memprediksi jumlah pendaftar haji di wilayah Kemenag Lhokseumawe menggunakan algoritma Random Forest. Dataset yang digunakan pada penelitian ini diambil dari Sistem Komputerisasi Haji Terpadu yang terdiri dari usia, jenis kelamin, status pekerjaan, tingkat pendidikan, status pernikahan, dan tren pendaftaran dari tahun ke tahun. Data kemudian melalui tahap pemrosesan seperti pengkodean variabel kategorik dan normalisasi fitur numerik, lalu data dibagi menjadi data latihan dan pengujian. Hasil akurasi tidak disebutkan pada penelitian ini tetapi dengan adanya penelitian ini, algoritma Random Forest terbukti dapat meningkatkan efisiensi perencanaan kuota dan pengelolaan sumber daya secara data [19].

Penelitian kesebelas dengan judul “Exploratory Study on Classification of Diabetes Mellitus Through a Combined Random Forest Classifier” yang berfokus pada peningkatan akurasi klasifikasi penyakit diabetes mellitus (DM) melalui penerapan algoritma Random Forest. Dataset yang digunakan terdiri dari 4.105 baris dan dibagi menjadi 70% untuk pelatihan dan 30% untuk pengujian, dengan proses pengujian diulang sebanyak 100 kali untuk memperoleh hasil rata-rata yang stabil. Hasil dari evaluasi model yang telah dilakukan menunjukkan bahwa algoritma Random Forest mendapatkan akurasi tertinggi pada angka 89% persen. Evaluasi model dilakukan dengan menggunakan metrik akurasi, presisi, *recall*, F1-score, dan AUC [6].

Penelitian kedua belas dengan judul “Penerapan Algoritma K-Nearest Neighbors dalam Sistem Rekomendasi Makanan Berdasarkan Kebutuhan Nutrisi dengan Content-Based Filtering” bertujuan untuk menerapkan algoritma *machine learning* KNN pada sistem rekomendasi makanan dengan teknik *content-based filtering*. Penelitian tersebut dilatar belakangi oleh pentingnya nutrisi yang cukup dalam pola makan harian, namun masih banyak individu yang masih awam dalam memilih makanan yang bernutrisi cukup. Kumpulan data yang digunakan untuk mengembangkan model *machine learning* diperoleh dari sebuah *website* yang menyediakan resep makanan beserta detail nutrisi dan kalori pada makanan tersebut, data diambil menggunakan metode *web scraping*. Algoritma KNN berperan untuk menentukan makanan terdekat berdasarkan kebutuhan pengguna, dengan menggunakan *Euclidean Distance* untuk mengukur kesamaan makanan berdasarkan nutrisi dan kalori makanan, proses tersebut menghasilkan 50 makanan terdekat dengan makanan pilihan pengguna. Dari 50 makanan tersebut dilakukan teknik *content-based filtering* untuk mencari beberapa data yang sesuai, dari hasil model *machine learning* tersebut dapat disimpulkan bahwa algoritma KNN dan CBF terbukti efektif dalam memberikan rekomendasi makanan yang sesuai dengan kebutuhan nutrisi individu [20].

Penelitian ketiga belas dengan judul “Pemanfaatan Metode Collaborative Filtering dengan Algoritma KNN pada Sistem Rekomendasi

Produk” bertujuan untuk membahas tentang pemanfaatan metode Item-based Collaborative Filtering dengan algoritma KNN dalam sistem rekomendasi produk *e-commerce*. Permasalahan yang diangkat mengenai pengguna *e-commerce* yang kesulitan dalam memilih produk yang ingin dibeli, sistem rekomendasi menjadi solusi dari permasalahan tersebut. Metode yang digunakan adalah metode Item-Based Collaborative Filtering dengan algoritma KNN untuk meningkatkan akurasi, data yang digunakan pada penelitian ini diambil dari Amazon *Product Review* dengan tahap pengolahan dimulai dari pengecekan statistik data, EDA, dan data *splitting*. Setelah data diolah, diterapkan permodelan dengan Item-Based Collaborative dan KNN dengan menentukan K dalam KNN lalu dihitung nilai kemiripan dengan *Cosine Similarity* dan menerapkan Item-Based Collaborative. Dari model yang telah dibuat, dievaluasi dengan menggunakan metrik *Mean Absolute Error* (MAE) yang mendapatkan nilai 1.05 dan metrik *Root Mean Square Error* (RMSE) yang mendapatkan nilai 1.36, dari nilai tersebut dapat disimpulkan bahwa metode Item-based Collaborative Filtering dengan KNN terbukti efektif dalam sistem rekomendasi produk [21].

Penelitian keempat belas dengan judul “Classification of Movie Recommendation on Netflix Using Random Forest Algorithm” bertujuan untuk mengembangkan sebuah sistem rekomendasi film dengan menggunakan algoritma *decision tree* dan Random Forest pada dataset Netflix. Dataset yang digunakan pada penelitian ini terdiri dari 5850 baris dengan 15 kolom seperti judul, jenis tayangan, deskripsi, tahun rilis, sertifikasi usia, durasi, genre, dan negara produksi. Hasil evaluasi dengan menggunakan algoritma Random Forest mencapai tingkat akurasi sebesar 96,92% yang meningkat dari 96,84% sebelum seleksi fitur. Sedangkan algoritma *decision tree* mencapai tingkat akurasi sebesar 96,07% yang meningkat dari 95,64% sebelum seleksi fitur. Evaluasi performa model dilakukan dengan menggunakan *precision metrix*, recall dan F1-score untuk memberikan gambaran yang menyeluruh [22].

Penelitian kelima belas dengan judul "Rancang Bangun Sistem Rekomendasi Makanan Alternatif Berkalori Lebih Rendah Berbasis Konten

Menggunakan Hierarchical Clustering" bertujuan untuk membahas pentingnya menjaga pola makan yang sehat untuk terhindar dari berbagai jenis penyakit, namun memilih makanan alternatif yang lebih sehat tidaklah mudah. Data yang digunakan pada penelitian ini diambil dari *website* FatSecret lalu data diproses menggunakan metode *Hierarchical Clustering* yang berguna untuk mengelompokkan makanan berdasarkan kemiripan bahan dan jumlah kalori. Evaluasi dilakukan melalui eksperimen terhadap 10 makanan yang diuji pada 32 responden dengan skala libert, hasil evaluasi mendapatkan indeks kepuasan sebesar 71.3% yang cukup memuaskan tetapi masih dapat dikembangkan karena beberapa rekomendasi makanan juga masih kurang sesuai dengan ekspektasi pengguna [23].

Untuk memperkuat posisi dan kontribusi penelitian ini, dilakukan telaah terhadap beberapa penelitian terdahulu yang memiliki topik serupa, khususnya yang berkaitan dengan klasifikasi data dan sistem rekomendasi. Perbandingan dilakukan untuk melihat keunggulan dan keterbatasan dari masing-masing penelitian. Hal ini bertujuan untuk menunjukkan gap atau celah penelitian yang masih dapat dikembangkan, sekaligus sebagai dasar penguatan terhadap pendekatan yang digunakan. Rangkuman hasil perbandingan tersebut dapat dilihat pada Tabel 2.1.

Tabel 2.1 Perbandingan penelitian terdahulu dengan penelitian penulis

	[14]	[13]	[20]	[19]	[18]	[16]	Peneliti
Topik	Metode Machine Learning untuk Klasifikasi Data Gizi Balita dengan Algoritma Naïve Bayes, KNN dan Decision Tree	Penerapan Algoritma k-Nearest Neighbors pada Aplikasi Rekomendasi Makanan Sehat Berbasis Mobile	Penerapan Algoritma K-Nearest Neighbors dalam Sistem Rekomendasi Makanan Berdasarkan Kebutuhan Nutrisi dengan Content-Based Filtering	Application of the Random Forest Algorithm for Predicting Hajj Registration Numbers at Kemenag Lhokseumawe	Healthcare Recommender System Using Random Forest Classifier	Random Forest Classification Method for Predicting Intertidal Wetland Migration Under Sea Level Rise	Klasifikasi Status Gizi dan Rekomendasi Pola Makan untuk Pencegahan Stunting Menggunakan Random Forest dan KNN
Tujuan	Klasifikasi status gizi	Rekomendasi makanan sehat	Rekomendasi makanan berdasarkan kebutuhan nutrisi	Prediksi jumlah pendaftar haji	Rekomendasi obat berdasarkan gejala	Prediksi migrasi lahan basah akibat SLR	Memprediksi status gizi dan memberikan rekomendasi makanan berdasarkan kondisi anak

	[14]	[13]	[20]	[19]	[18]	[16]	Peneliti
Algoritma	NB, KNN, <i>Decision Tree</i>	KNN (cosine)	KNN, CBF (Euclidean)	Random Forest	Random Forest	Random Forest	KNN (Euclidean), Random Forest
Sumber dataset	Data posyandu	FatSecret	<i>Website</i> resep makanan	SISKOHAT	Tidak diketahui	DEM, peta vegetasi makrofit	Scraping resep & biodata anak
Metode Evaluasi	Tidak diketahui	Cosine distance	Euclidean distance	Tidak diketahui	Tidak diketahui	Perbandingan hasil prediksi dengan kondisi aktual	Confusion matrix & kecocokan AKG
Terdapat sistem yang diintegrasikan	Tidak ada	Ada	Ada	Tidak	Ada	Tidak ada	Ada
Akurasi diatas 90%	Tidak	Tidak diketahui	Tidak diketahui	Tidak diketahui	Ya	Tidak diketahui	Ya