

BAB II

TINJAUAN PUSTAKA

Tinjauan pustaka merupakan bagian yang membahas berbagai referensi dan penelitian terdahulu yang relevan dengan pengembangan sistem rekomendasi lowongan kerja berbasis web. Bagian ini menyajikan konsep, teori, serta studi-studi sebelumnya yang menjadi dasar perancangan dan implementasi sistem dalam penelitian ini.

Penelitian oleh Sharma et al. mengusulkan *Smart Job Recommendation System* yang menggabungkan *content-based* dan *collaborative filtering*. Dataset yang digunakan mencakup atribut seperti pendidikan, pengalaman kerja, dan keterampilan. Sistem ini berhasil mengatasi masalah *cold-start* dan *skill-based matching* dengan pendekatan *hybrid*. Hasil evaluasi menunjukkan bahwa sistem mencapai akurasi sebesar 87%, *precision* sebesar 0.84, *recall* sebesar 0.83, dan F1-score sebesar 0.835. Performa ini menunjukkan bahwa pendekatan *hybrid* memberikan hasil yang lebih baik dibandingkan dengan metode tunggal [5].

Penelitian oleh Karadkar mengusulkan Job Recommendation System menggunakan beberapa algoritma machine learning seperti Logistic Regression, K-Nearest Neighbor (KNN), Random Forest, Support Vector Machine (SVM), dan Naive Bayes. Dataset mencakup profil pengguna dan lowongan pekerjaan yang diproses menggunakan TF-IDF vectorization. Model Random Forest terbukti memberikan performa terbaik dengan akurasi sebesar 88,7% dan F1-Score sebesar 87,0%. Precision dan recall dari Random Forest juga menunjukkan hasil yang konsisten tinggi. Model dievaluasi menggunakan 5-fold cross-validation dan kurva ROC untuk memastikan robustness dari sistem rekomendasi [6].

Penelitian oleh Pratama et al. dari Universitas Islam Indonesia menerapkan Model Klasifikasi Calon Mahasiswa Baru untuk Sistem Rekomendasi Program Studi Sarjana Berbasis *Machine Learning*. Dataset terdiri dari data pribadi dan akademik semua mahasiswa dan alumni program sarjana di Universitas Islam Indonesia (UII). Tiga algoritma machine learning dibandingkan: *Random Forest* (RF), *Multinomial Logistic Regression* (MLR), dan *Support Vector Machine* (SVM). Hasil penelitian menunjukkan bahwa model *Random Forest* memberikan performa terbaik dengan akurasi sebesar 86%, precision sebesar 84%, recall sebesar 86%, dan AUC sebesar 97%. Model *Random Forest* terbukti memiliki kinerja yang jauh lebih baik dibandingkan dengan MLR maupun SVM [7].

Penelitian oleh Susanti et al. membandingkan lima algoritma *supervised learning*, yaitu *Decision Tree* (DT), *Random Forest* (RF), *Logistic Regression* (LR), *Light Gradient Boosting Machine* (LGBM), dan *Support Vector Machine* (SVM), dalam klasifikasi prediksi *dropout* mahasiswa. Dataset terdiri dari data mahasiswa Universitas Sebelas Maret tahun 2013 dengan total 2.476 sampel. Berdasarkan hasil eksperimen dengan menggunakan teknik *preprocessing*, *resampling SMOTE*, dan *K-Fold Cross-Validation*, *Random Forest* memiliki tingkat akurasi tertinggi sebesar 96,01%, diikuti oleh LGBM (95,26%), *Decision Tree* (91,24%), *Logistic Regression* (83,68%), dan SVM (83,19%). Sedangkan *Logistic Regression* unggul dari sisi kecepatan dan efisiensi komputasi pada dataset berskala besar [8].

Penelitian oleh Probst et al. fokus pada perbandingan komprehensif antara *Random Forest* dan *Logistic Regression* menggunakan 243 dataset riil. Studi ini menunjukkan bahwa *Random Forest* memberikan performa lebih baik dibandingkan *Logistic Regression* pada sekitar 69% dari dataset yang diuji. Perbedaan rata-rata antara RF dan LR adalah 0.029 (95%-CI = [0.022, 0.038]) untuk akurasi, 0.041 (95%-CI = [0.031, 0.053]) untuk *Area Under the Curve*

(AUC), dan -0.027 (95%-CI = [-0.034, -0.021]) untuk Brier score. Penelitian ini juga menunjukkan bahwa proses tuning parameter secara sistematis mampu meningkatkan performa model secara signifikan [9].

Penelitian oleh H. A. El-Keiey mengembangkan sistem rekomendasi karier bagi mahasiswa sarjana dengan menggunakan berbagai algoritma *machine learning* dan metode seleksi fitur. Dataset terdiri atas data akademik dan minat karier mahasiswa. Hasil penelitian menunjukkan bahwa seleksi fitur yang tepat dapat meningkatkan akurasi prediksi rekomendasi karier secara signifikan [10].

Penelitian oleh B. Betrand mengusulkan sistem rekomendasi karier berbasis *Decision Tree*, *Random Forest*, dan *Naïve Bayes*. Berdasarkan hasil evaluasi, *Random Forest* menghasilkan akurasi tertinggi sebesar 92,5%, diikuti oleh *Decision Tree* dan *Naïve Bayes*. Sistem ini juga dilengkapi fitur *career guidance* yang menampilkan rekomendasi pekerjaan sesuai latar belakang pengguna [11].

Berdasarkan penelitian-penelitian terdahulu tersebut, dapat disimpulkan bahwa penerapan **supervised learning** dalam sistem rekomendasi pekerjaan telah banyak dilakukan dengan menggunakan berbagai algoritma, seperti *Decision Tree*, *Random Forest*, dan *Logistic Regression*. Masing-masing algoritma memiliki kelebihan dan kelemahan, baik dari segi interpretabilitas, kompleksitas komputasi, maupun akurasi. Penelitian ini berkontribusi dengan membandingkan performa ketiga algoritma tersebut, memilih model terbaik melalui optimasi *hyperparameter*, serta mengintegrasikannya ke dalam sistem berbasis web menggunakan **FastAPI** sebagai *backend API* dan **Vue.js** sebagai *frontend framework* untuk menampilkan hasil rekomendasi lowongan kerja secara *real-time*.

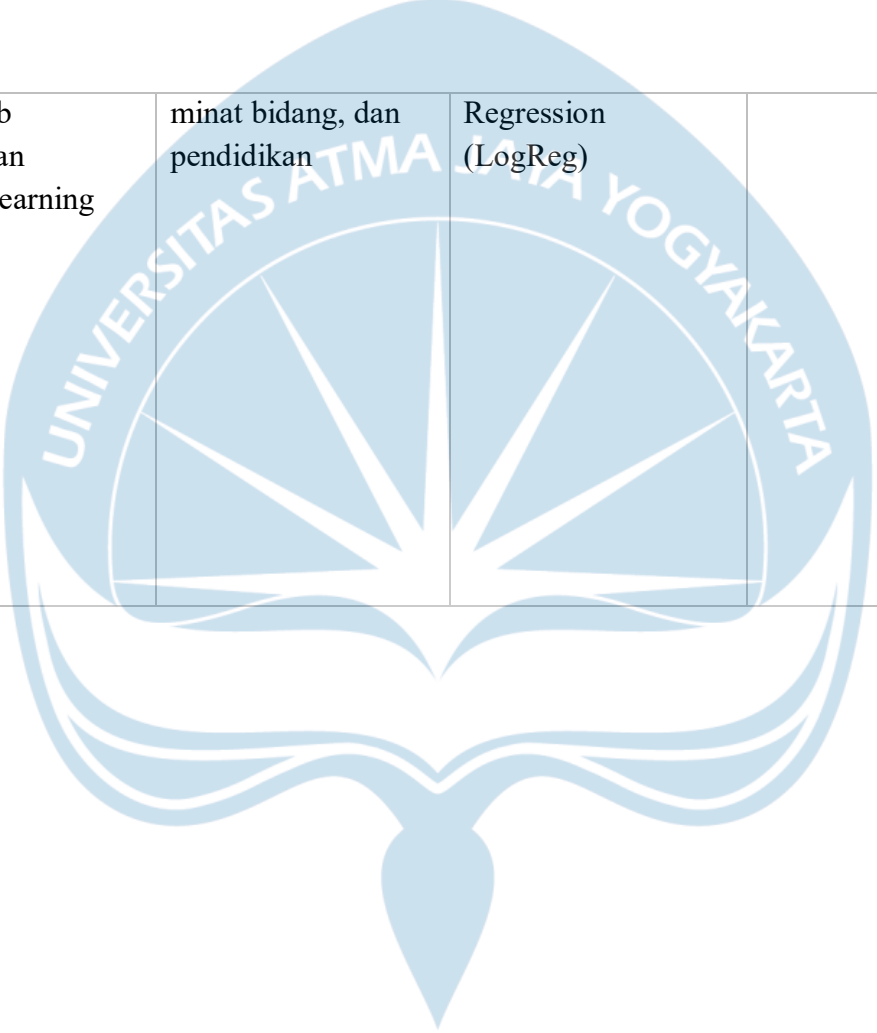
Tabel 2. 1 Perbandingan Penelitian

No	Penelitian	Tujuan Penelitian	Dataset dan Sumber	Model yang Dibandingkan	Model Terbaik	Fitur Rekomendasi yang Dikembangkan
1	Sharma et al.[5]	Mengembangkan Smart Job Recommendation System dengan pendekatan hybrid untuk mengatasi masalah cold-start dan skill-based matching	Dataset mencakup atribut pendidikan, pengalaman kerja, dan keterampilan pengguna	Content-Based Filtering (CBF), Collaborative Filtering (CF), Hybrid Model	Hybrid Model (Akurasi: 87%, Precision: 0.84, Recall: 0.83, F1-Score: 0.835)	- Pencocokan keterampilan (skill matching) - Pendekatan hybrid untuk mengatasi cold-start problem - Rekomendasi berbasis profil pengguna
2	Karadkar[6]	Mengembangkan Job Recommendation System menggunakan multiple machine learning algorithms	Dataset profil pengguna dan lowongan pekerjaan yang diproses dengan TF-IDF vectorization	Logistic Regression, K-Nearest Neighbor (KNN), Random Forest, Support Vector Machine	Random Forest (Akurasi: 88.7%, F1-Score: 87.0%)	- TF-IDF vectorization untuk text processing - 5-fold cross-validation

		dengan TF-IDF vectorization		(SVM), Naive Bayes		- Kurva ROC untuk evaluasi robustness - Pencocokan profil pengguna dengan lowongan kerja
3	Pratama et al. (Universitas Islam Indonesia)[7]	Mengembangkan Model Klasifikasi Calon Mahasiswa Baru untuk Sistem Rekomendasi Program Studi Sarjana	Data pribadi dan akademik mahasiswa dan alumni program sarjana Universitas Islam Indonesia (UII)	Random Forest (RF), Multinomial Logistic Regression (MLR), Support Vector Machine (SVM)	Random Forest (Akurasi: 86%, Precision: 84%, Recall: 86%, AUC: 97%)	- Rekomendasi program studi berbasis data akademik - Klasifikasi calon mahasiswa baru - Sistem rekomendasi untuk pemilihan jurusan
4	Susanti et al.[8]	Membandingkan algoritma supervised learning untuk prediksi dropout mahasiswa	Data mahasiswa Universitas Sebelas Maret tahun 2013 (total 2.476 sampel) dengan teknik preprocessing, resampling	Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), Light Gradient Boosting Machine (LGBM), Support	Random Forest (Akurasi: 96.01%), diikuti LGBM (95.26%), DT (91.24%), LR	- Prediksi dropout mahasiswa - SMOTE untuk handling imbalanced data

			SMOTE, dan K-Fold Cross-Validation	Vector Machine (SVM)	(83.68%), SVM (83.19%)	- K-Fold Cross-Validation untuk validasi model - Early warning system untuk mahasiswa berisiko dropout
5	Probst et al.[9]	Melakukan perbandingan komprehensif antara Random Forest dan Logistic Regression pada large-scale benchmark	243 dataset riil dari berbagai domain	Random Forest (RF), Logistic Regression (LR)	Random Forest (Unggul di 69% dataset, Δ Akurasi: 0.029, Δ AUC: 0.041)	- Systematic hyperparameter tuning - Benchmark comparison framework - Performance evaluation across multiple metrics - Guidelines untuk pemilihan model
6	El-Keiey et al.[9]	Mengembangkan sistem rekomendasi karier untuk mahasiswa sarjana	Data akademik dan minat karier mahasiswa	Berbagai algoritma machine learning dengan metode seleksi fitur	Model dengan feature selection yang tepat (peningkatan	- Feature selection untuk meningkatkan akurasi

		menggunakan feature selection			akurasi signifikan)	- Rekomendasi karier berbasis data akademik - Sistem rekomendasi untuk mahasiswa sarjana
7	Betrand et al.[10]	Mengembangkan sistem rekomendasi karier dengan career guidance feature	Dataset karier dan latar belakang pengguna	Decision Tree, Random Forest, Naïve Bayes	Random Forest (Akurasi: 92.5%), diikuti Decision Tree dan Naïve Bayes	- Career guidance system - Rekomendasi pekerjaan berbasis latar belakang pengguna - Interface untuk konsultasi karier - Sistem klasifikasi multi-class untuk berbagai jalur karier
8	Fidelis Vendriko Gerardy	Mengembangkan sistem rekomendasi lowongan kerja	Data CV pelamar kerja yang mencakup skill,	Random Forest (RF), Decision Tree (DT), Logistic	- (Data hasil belum tersedia)	- Web-based interface



		berbasis web menggunakan supervised learning	minat bidang, dan pendidikan	Regression (LogReg)		- Form input untuk skill, pendidikan, dan bahasa pemrograman - Analisis dan rekomendasi pekerjaan otomatis - Sistem matching berbasis profil pelamar
--	--	--	---------------------------------	------------------------	--	--

Berdasarkan Tabel Perbandingan Penelitian, dapat dilihat bahwa penelitian terdahulu berfokus pada penerapan algoritma *machine learning* dalam sistem rekomendasi pekerjaan dan karir dengan pendekatan yang beragam.

Beberapa penelitian seperti Pratama et al., Susanti et al., dan Betrand et al. menekankan pada penggunaan algoritma *Decision Tree*, *Random Forest*, dan *Logistic Regression* untuk melakukan klasifikasi data pelamar kerja dan mahasiswa. Hasil penelitian tersebut menunjukkan bahwa Random Forest memiliki akurasi yang lebih tinggi dan lebih stabil dibandingkan model lainnya karena menggunakan agregasi dari banyak pohon keputusan (ensemble method). Pratama et al. mencatat bahwa Random Forest mencapai akurasi 86% dengan AUC 97%, sementara Susanti et al. melaporkan akurasi hingga 96,01%, dan Betrand et al. bahkan mencatat akurasi 92,5%. Konsistensi performa tinggi ini membuktikan keunggulan ensemble method dalam sistem rekomendasi.[5]

Sementara itu, penelitian Probst et al. berfokus pada perbandingan komprehensif antara Random Forest dan Logistic Regression menggunakan 243 dataset riil, dengan tujuan memberikan benchmark yang objektif. Penelitian ini menegaskan bahwa Random Forest unggul di 69% dataset yang diuji dengan perbedaan rata-rata akurasi 0,029 dan AUC 0,041. Studi ini juga menegaskan pentingnya proses hyperparameter tuning dalam meningkatkan hasil prediksi model dan menurunkan tingkat kesalahan klasifikasi.[6]

Penelitian Sharma et al. mengembangkan Smart Job Recommendation System dengan pendekatan hybrid yang menggabungkan Content-Based Filtering (CBF) dan Collaborative Filtering (CF). Model hybrid yang diusulkan berhasil mengatasi masalah cold-start dan skill-based matching, dengan hasil menunjukkan akurasi 87%, precision 0,84, dan recall 0,83. Pendekatan ini menunjukkan efektivitas kombinasi dua teknik rekomendasi dalam memberikan hasil yang lebih akurat dibandingkan penggunaan metode tunggal.[7]

Kemudian, Karadkar mengusulkan Job Recommendation System menggunakan multiple machine learning algorithms dengan TF-IDF vectorization. Hasil penelitian menunjukkan bahwa Random Forest memberikan performa terbaik dengan akurasi 88,7% dan F1-Score 87,0%. Model dievaluasi menggunakan 5-fold cross-validation dan kurva ROC untuk memastikan robustness sistem. Penelitian ini menunjukkan pentingnya text processing yang baik dalam meningkatkan kualitas rekomendasi pekerjaan.[8]

Selanjutnya, penelitian El-Keiey et al. menyoroti penerapan feature selection dalam sistem rekomendasi karier untuk mahasiswa sarjana. Hasil penelitian menunjukkan bahwa pemilihan fitur yang tepat dapat meningkatkan akurasi prediksi rekomendasi karier secara signifikan. Penelitian ini menekankan bahwa tidak semua fitur memberikan kontribusi yang sama dalam proses klasifikasi, sehingga seleksi fitur menjadi tahap penting dalam pengembangan model yang efisien dan akurat.

Sementara itu, penelitian Betrand et al. mengembangkan career guidance system yang tidak hanya memberikan rekomendasi, tetapi juga menyediakan fitur konsultasi karier interaktif. Dengan menggunakan Decision Tree, Random Forest, dan Naïve Bayes, sistem ini mampu menghasilkan rekomendasi jalur karier berdasarkan latar belakang pengguna. Random Forest kembali menunjukkan performa terbaik dengan akurasi 92,5%, membuktikan konsistensinya sebagai algoritma pilihan dalam domain rekomendasi karier.[9]

Dari berbagai penelitian tersebut, dapat disimpulkan bahwa sebagian besar penelitian terdahulu berfokus pada peningkatan akurasi sistem rekomendasi melalui pendekatan model tertentu, seperti *Random Forest*, *Logistic Regression*, atau *embedding-based models*. Namun, penelitian-penelitian tersebut belum secara eksplisit membandingkan performa beberapa algoritma *supervised learning* dalam satu studi yang sama, serta belum banyak yang mengintegrasikan model tersebut ke dalam sistem web berbasis API yang interaktif.[10]

Dengan demikian, penelitian ini memiliki keunikan dengan melatih dan membandingkan tiga algoritma *supervised learning* yaitu Decision Tree (DT), Random Forest (RF), dan Logistic Regression (LR) pada dataset profil pelamar kerja bidang IT, lalu memilih model terbaik berdasarkan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score. Selain itu, model terbaik akan diintegrasikan ke dalam sistem berbasis web menggunakan **FastAPI** sebagai *backend* dan **Vue.js** sebagai *frontend*, agar hasil rekomendasi pekerjaan dapat diakses secara real-time dan interaktif oleh pengguna.

