



***HYBRID CLUSTERING DENGAN K-MEANS DAN RANDOM
FOREST UNTUK PREDIKSI KUALITAS PERGURUAN TINGGI***

SEMINAR BIDANG KAJIAN

DAVID AULIA AKBAR ADHIEPUTRA

99223104

ANGKATAN 31 S3 TI

PROGRAM DOKTOR TEKNOLOGI INFORMASI

UNIVERSITAS GUNADARMA

JUNI, 2023

DAFTAR ISI

	Halaman
DAFTAR ISI	ii
1 PENDAHULUAN	3
1.1 Latar Belakang.....	3
1.2 Batasan dan Tujuan	5
1.3 Kontribusi	5
2 TINJAUAN PUSTAKA	5
2.1 Tinjauan Penelitian Terkait	5
2.2 Perbandingan Tinjauan penelitian	15
• Random Forest:	15
• K-means:.....	15
3 METODE PENELITIAN.....	16
3.1 Motivasi.....	16
3.2 Framework Riset.....	17
3.3 Pendekatan.....	19
DAFTAR PUSTAKA.....	21

1 PENDAHULUAN

1.1 Latar Belakang

Teknologi Informasi (TI) telah mengalami perkembangan pesat dan penerapannya telah menyentuh berbagai bidang ilmu, industri, dan pemerintahan. Di sektor industri, TI telah membantu dalam otomatisasi proses produksi, manajemen rantai pasokan, serta analisis data besar untuk pengambilan keputusan yang lebih baik. Di bidang ilmu pengetahuan, TI mendukung penelitian melalui komputasi yang lebih cepat dan analisis data yang lebih akurat. Dalam pemerintahan, penerapan TI sangat vital untuk meningkatkan efisiensi dan transparansi. Sistem e-government, misalnya, telah memudahkan akses masyarakat terhadap layanan publik dan mengurangi birokrasi. Dalam konteks pendidikan tinggi, teknologi ini dapat digunakan untuk mengelola data kelembagaan dan aktivitas akademik, serta mendukung sistem penjaminan mutu.

Sistem penjaminan mutu di perguruan tinggi merupakan elemen penting untuk memastikan bahwa lembaga pendidikan tinggi mampu memenuhi standar yang telah ditetapkan. Pemangku kepentingan, termasuk pemerintah, institusi pendidikan, dan masyarakat, membutuhkan mekanisme yang efektif untuk menilai dan meningkatkan kualitas pendidikan. Dalam hal ini, penelitian yang berfokus pada penerapan metode hybrid clustering dengan K-Means dan Random Forest untuk prediksi kualitas perguruan tinggi menjadi sangat relevan. Penelitian ini tidak hanya membantu mengidentifikasi dan mengelompokkan institusi berdasarkan indikator kinerja, tetapi juga memungkinkan pemangku kepentingan untuk membuat keputusan yang lebih informatif dan strategis untuk perbaikan kualitas pendidikan.

Penelitian terdahulu telah menunjukkan berbagai pendekatan dan metode untuk mengatasi masalah penjaminan mutu pendidikan tinggi. Beberapa penelitian menawarkan pendekatan inovatif menggunakan metode machine learning. Bicego (2019) mengembangkan K-Random Forests yang memanfaatkan setiap forest untuk memodelkan satu cluster, diperbarui secara iteratif mirip dengan K-means, memberikan fleksibilitas dalam deskripsi cluster. Deepika (2019) mengintegrasikan metode seleksi fitur Relief-F dengan Random Forest untuk meningkatkan akurasi

prediksi kinerja akademik mahasiswa. (Dai, 2022) meningkatkan algoritma Random Forest dengan menggabungkan XGBoost dan GBDT, menunjukkan peningkatan dalam akurasi dan stabilitas model. (Wiyono *et al.*, 2019) menggunakan seleksi fitur dengan RandomForest untuk memprediksi status aktif mahasiswa, efektif dalam mengidentifikasi fitur penting. (Liang *et al.*, 2019) menggabungkan Random Forest dan KNN, meningkatkan akurasi klasifikasi dengan memilih dan menghapus fitur berdasarkan indeks Gini. Pendekatan-pendekatan ini menunjukkan berbagai cara dalam menggabungkan teknik-teknik machine learning untuk analisis data pendidikan yang kompleks.

Analisis kelemahan dari penelitian-penelitian terkait metode Random Forest dan K-means menunjukkan beberapa gap yang dapat dijadikan dasar untuk mengusulkan topik "Hybrid Clustering University Quality Prediction" dalam penelitian disertasi. Misalnya, meskipun (Bicego, 2019) menunjukkan keefektifan K-Random Forests dalam clustering, metode ini memerlukan tuning parameter yang cermat dan mungkin tidak optimal tanpa konfigurasi awal yang tepat. (Deepika and Sathyanarayana, 2019) berhasil mengintegrasikan seleksi fitur dengan Random Forest untuk prediksi kinerja akademik, namun kompleksitas dan waktu komputasi menjadi tantangan. (Dai, 2022) menggabungkan beberapa algoritma untuk meningkatkan prediksi, namun kompleksitas konfigurasi juga menjadi kekurangan.

Dari analisis kebutuhan, terlihat bahwa ada kebutuhan untuk mengembangkan metode yang dapat secara efisien mengintegrasikan teknik clustering dan prediksi untuk menilai kualitas perguruan tinggi dengan lebih akurat dan efisien. Oleh karena itu, topik "Hybrid Clustering University Quality Prediction" diusulkan untuk mengatasi kekurangan ini dengan mengembangkan model hybrid yang menggabungkan kelebihan dari K-means dan Random Forest. Tujuan umum dari penelitian ini adalah untuk menciptakan model yang dapat lebih akurat dalam mengklasifikasikan dan memprediksi kualitas perguruan tinggi berdasarkan berbagai indikator, sambil mengurangi kompleksitas dan waktu komputasi. Model ini diharapkan dapat memberikan wawasan yang lebih mendalam dan bermanfaat bagi pengambil keputusan di sektor pendidikan untuk meningkatkan standar dan kualitas pendidikan tinggi.

1.2 Batasan dan Tujuan

Penelitian ini memiliki beberapa batasan, yaitu: keterbatasan data yang hanya berasal dari PDDikti tanpa mengintegrasikan data eksternal seperti peringkat internasional atau data alumni, keterbatasan dalam evaluasi model yang mungkin tidak mencakup semua konteks geografis dan temporal, serta potensi kesulitan dalam interpretasi hasil yang dihasilkan oleh algoritma Random Forest. Tujuan utama penelitian ini adalah untuk mengembangkan model hybrid yang menggabungkan K-Means dan Random Forest guna meningkatkan akurasi prediksi kualitas perguruan tinggi, dengan memanfaatkan data kelembagaan dan aktivitas akademik. Penelitian ini juga bertujuan untuk memberikan model yang dapat diinterpretasikan dengan baik oleh pemangku kepentingan, serta memastikan model ini dapat diterapkan secara luas di berbagai konteks untuk mendukung penjaminan mutu pendidikan tinggi.

1.3 Kontribusi

Kontribusi dari penelitian ini adalah:

1. Penelitian ini mengembangkan model hybrid yang meningkatkan akurasi prediksi kualitas perguruan tinggi dengan menggabungkan metode K-Means dan Random Forest, memperkaya literatur data science dalam pendidikan tinggi.
2. Penelitian ini menyediakan alat analisis yang lebih akurat dan dapat diinterpretasikan untuk Dirjen Pendidikan Tinggi, membantu dalam menetapkan kebijakan penjaminan mutu pendidikan yang lebih efektif dan efisien.

2 TINJAUAN PUSTAKA

Bab ini merupakan studi literatur tentang materi – materi yang berhubungan dengan proses penelitian.

2.1 Tinjauan Penelitian Terkait

Berikut adalah penelitian-penelitian yang telah dilakukan terkait dengan perkembangan model klasterisasi dengan pendekatan algoritma unsupervised learning maupun model klasifikasi dengan pendekatan algoritma supervised learning.

Tabel 2.1 Studi Penelitian yang Pernah Dilakukan

No	Authors	Title	Research Question	Metodology	Algorithm	Main Findings
1	(Ramadhan, Susetyo and Indahwati, 2021)	Classification Modelling of Random Forest to Identify the Important Factors in Improving the Quality of Education	What are the key factors that influence the quality of education at the high school level, as identified through random forest classification modeling of national exam scores and accreditation results?	The study used random forest for multi-class classification to analyze UNBK scores and accreditation results in order to identify important factors for improving the quality of education.	Random forest	- The study used random forest classification modeling to identify the important factors in improving the quality of education at the high school level, based on national exam scores and accreditation results. - The model achieved high classification accuracy and identified 15 specific accreditation instrument items as important factors, which were related to three key national education standards: school facilities, teacher and education staff, and graduate competency. - The study recommends that regional governments and education units collaborate to improve these three key standards in order to improve education quality.
2	(Zhang <i>et al.</i> , 2021)	A Modified Random Forest Based on Kappa Measure and Binary Artificial Bee Colony Algorithm	How can we select high-quality decision trees from a random forest to improve the overall accuracy of the model?	The methodology involves using Kappa measure for pre-pruning to eliminate low-quality decision trees, and then using an improved binary artificial bee colony (IBABC) algorithm for secondary pruning	Kappa measure and improved binary artificial bee colony (IBABC) algorithm	- The proposed Kappa+IBABC method improves the performance of random forest on most datasets, using fewer decision trees. - The Kappa+IBABC method is used to predict haze in China and reduce the impact of haze on people's daily life.

				to select the higher-performing decision trees for the final voting in the random forest.		
3	(Bicego, 2019)	K-Random Forests: a K-means style algorithm for Random Forest clustering	How can Random Forests be used for clustering?	1. - The methodology uses multiple Random Forests, with each one modeling a single cluster - The Random Forests are iteratively updated using a K-means-like clustering algorithm	K-Random Forests (K-RF)	<ul style="list-style-type: none"> - The authors present a novel clustering approach called K-Random Forests (K-RF) that uses multiple Random Forests, each modeling one cluster, and iteratively updates these Random Forests using a K-means-like algorithm. - The authors evaluated this K-RF approach on 5 datasets and found that it represents a valid alternative to other Random Forest clustering algorithms as well as other established clustering approaches.
4	(Deepika and Sathyanarayana, 2019)	Relief-F and Budget Tree Random Forest Based Feature Selection for Student Academic Performance Prediction	How can we develop an effective feature selection method to improve the prediction of student academic performance by reducing redundant and irrelevant features in large educational datasets?	The methodology used in the study is a feature selection method called RFBT-RF (Relief-F and Budget Tree Random Forest) that selects relevant features and removes irrelevant/redundant features to improve the performance of classifiers used for predicting student academic performance.	The specific algorithms introduced, studied, or used in the study are: Relief-F, Budget Tree Random Forest (BT-RF), and the hybrid RFBT-RF feature selection method, which was applied to multiple	<ul style="list-style-type: none"> - The proposed RFBT-RF feature selection method selects only relevant features and improves the performance of multiple classifiers (DT, NB, SVM, KNN) in predicting student academic performance. - The RFBT-RF method achieved 6.85% higher accuracy compared to the existing Logistic Regression (LR) model. - The RFBT-RF method achieved 97.88% prediction accuracy on the collected school database using the Decision Tree (DT) classifier, outperforming the LR method.

					machine learning classifiers including Decision Tree (DT), Naive Bayes (NB), Support Vector Machine (SVM), and K-Nearest Neighbour (KNN) for predicting student academic performance.	
5	(Prasetyowati, Maulidevi and Surendro, 2020)	Feature selection to increase the random forest method performance on high dimensional data	How can feature selection be used to improve the performance (execution time and accuracy) of the Random Forest method on high-dimensional datasets?	The methodology used in this study was feature selection using the Correlation-Based Feature Selection method with the BestFirst search algorithm, and testing was done 30 times using 10-fold cross-validation and a 70/30 train/test split.	Random Forest and Correlation-Based Feature Selection with BestFirst	- Using the Correlation-based feature selector (CfsSubsetEval) with the BestFirst method for feature selection can speed up the classification process of the Random Forest method on high-dimensional datasets and improve its accuracy, except on the CNAE-9 dataset where accuracy decreased, likely due to a "sparse problem".
6	(Dai, 2022)	Improving Random Forest Algorithm for University Academic Affairs	How can we use data from a network teaching platform to predict and	The methodology used in the study was: - Combining the	Random forest, XGBoost, and GBDT	- The proposed model based on an improved random forest algorithm achieves good prediction accuracy for

		Management System Platform Construction	analyze college students' academic performance and achievements in order to assist teachers in formulating effective teaching strategies?	advantages of random forest, XGBoost, and GBDT algorithms to create an improved random forest algorithm model - Testing the improved random forest algorithm model on a real dataset provided by a municipal education bureau - Demonstrating that the proposed model achieved good prediction accuracy and solved the stability problem of adding new data		college students' learning performance. - The model solves the stability problem of adding new data, which contributes to iterative optimization and improves the model's universality. - The model helps continuously track students' learning behavior characteristics across different semesters.
7	(Ghosh and Janan, 2021)	Prediction of Student's Performance Using Random Forest Classifier	What are the reasons for students' different types of results (performance) and how can we predict students' performance based on those factors?	- Used random forests (RF), a supervised machine learning algorithm, to classify and predict student performance - Obtained the input dataset by merging values from two surveys of students and experts, using fuzzy ANFIS analysis	Random Forest Classifier	- The random forest classifier was able to classify students' performance into different levels with 96.88% accuracy. - The proposed model can be used to predict course-wise student performance, and its accuracy can be further improved by adding more factors.
8	(Liang <i>et al.</i> , 2019)	An Improved Algorithm based on KNN and Random Forest	How can we develop an improved algorithm that combines KNN and random forest to	- Sorting features by importance using Gini index and random forest - Deleting unimportant	RFDKNN (an algorithm based on enhanced	- The RFDKNN algorithm, which combines an enhanced KNN and random forest, can achieve relatively high classification accuracy compared to

			achieve higher classification accuracy, especially on large datasets?	features based on the feature importance ranking, using a proportion r - Using an enhanced KNN algorithm that dynamically selects the optimal number of nearest neighbors and distance function	KNN and random forest)	other algorithms, especially on large datasets. - RFDKNN with $r=0.7$ (a parameter controlling feature selection) performs well. - RFDKNN outperforms other algorithms like Naive Bayes, Adaboost, Random Forest, and other KNN variants on most datasets, particularly large ones.
6	(Yi <i>et al.</i> , 2019)	A Novel Random Forest and its Application on Classification of Air Quality	How can we develop better methods for classifying air quality using machine learning, especially when the data is imbalanced?	- Proposed a new random forest method based on "samples grouped bootstrap" to handle imbalanced air quality data - Designed three experiments to evaluate the performance of this new method	A modified random forest algorithm using a samples grouped bootstrap approach	- The proposed random forest method based on samples grouped bootstrap improves upon standard random forest, especially when dealing with imbalanced air quality datasets. - The new method is much better at classifying minority samples in imbalanced datasets compared to standard random forest.
7	(Beaulac and Rosenthal, 2019)	Predicting University Students' Academic Success and Major Using Random Forests	1) Can an algorithm accurately predict whether a student will complete their undergraduate program? 2) Can an algorithm accurately predict the major that a student will complete, for those who	The study used classification trees and random forests as the main methodological approaches. Classification trees recursively partition the feature space to classify	Classification trees and random forests	- Random forests outperformed linear logistic models in predicting student outcomes. - Grades in low-grading departments were important predictors of whether a student would complete their program. - Allowing for students to have multiple majors could improve the accuracy of predicting a student's major.

			do complete their program?	<p>observations, while random forests improve on this by growing multiple trees on bootstrap samples and randomly selecting a subset of predictors at each split. The study also used two types of variable importance measures: Gini decrease importance, which measures the total decrease in impurity caused by a predictor across all trees, and permutation decrease importance, which measures the decrease in prediction accuracy when a predictor's values are randomly permuted.</p>		
8	(Wiyono <i>et al.</i> , 2019)	Feature Selection With the Random Forest Packages	What features can be used to predict which	- The methodology used feature selection with the	RandomForest	- The strongest features for predicting student active status are, in order: grade point (GP), grade point average (GPA),

		to Predict Student Performance	students will be inactive in the next semester?	RandomForest package in R to predict student active status. - The 7 features used were: - 1 output feature: active status of students - 6 input features: grade point (GP), grade point average (GPA), parent work, school majors, school category, and student hometown - The results showed the relative importance of the 6 input features, with GP and GPA being the strongest, followed by parent work, school majors, school category, and student hometown.		parent work, majors of origin, schools of origin, and student hometown.
	(Zhang and Bu, 2019)	Weighted KNN Algorithm Based on Random Forests	How can we develop a weighted KNN algorithm that overcomes the limitations of standard KNN in classifying	- Proposed a weighted KNN algorithm based on random forests - Used random forests to determine	Weighted KNN algorithm based on random forests	- The authors proposed a new weighted KNN algorithm based on random forests that overcomes the shortcomings of standard KNN in classifying unbalanced datasets and datasets with features of varying importance. - The new

			unbalanced datasets and datasets with features of varying importance?	the importance of each feature - Weighted the KNN algorithm based on the feature importance to overcome issues with unbalanced datasets and datasets with features of different importance		algorithm effectively improves the classification accuracy of KNN, as verified through experiments.
--	--	--	---	--	--	---

2.2 Perbandingan Tinjauan penelitian

Dalam memilih metodologi untuk penelitian tentang prediksi kualitas perguruan tinggi, K-means dan Random Forest dapat menjadi pilihan yang tepat berdasarkan analisis perbandingan penelitian-penelitian di atas. Berikut adalah analisis kelebihan dan kekurangan dari metodologi tersebut berdasarkan temuan utama dari beberapa penelitian:

- **Random Forest:**

Kelebihan: Random Forest terbukti efektif dalam menangani dataset dengan dimensi tinggi dan variabel yang banyak, seperti yang ditunjukkan dalam penelitian oleh (Prasetyowati, Maulidevi and Surendro, 2020) yang menunjukkan peningkatan kecepatan dan akurasi dengan seleksi fitur pada data berdimensi tinggi. (Ghosh and Janan, 2021) juga menunjukkan efektivitas Random Forest dalam prediksi kinerja mahasiswa dengan akurasi tinggi.

Kekurangan: Salah satu kelemahan adalah beberapa pohon keputusan berkualitas rendah dapat mengurangi akurasi keseluruhan dari Random Forest, seperti yang diidentifikasi oleh (Zhang and Bu, 2019) yang mengusulkan modifikasi untuk meningkatkan kualitas pohon keputusan.

- **K-means:**

Kelebihan: K-means digunakan dalam pendekatan clustering yang inovatif dengan Random Forest dalam penelitian oleh (Bicego, 2019), yang menunjukkan bahwa kombinasi K-means dengan Random Forest dapat memberikan deskriptor cluster yang fleksibel dan efektif.

Kekurangan: Metode ini mungkin tidak selalu efektif jika data tidak terdistribusi secara normal atau jika jumlah kluster tidak diketahui sebelumnya, yang bisa membatasi aplikasinya dalam beberapa konteks data.

Dengan mempertimbangkan kelebihan dari kedua metodologi ini, penggunaan K-means dan Random Forest dalam penelitian tentang prediksi kualitas perguruan tinggi dapat dijustifikasi. K-means dapat membantu dalam mengidentifikasi kelompok atau segmen dalam data yang berkaitan dengan kualitas perguruan tinggi, sementara Random Forest dapat digunakan untuk mengklasifikasikan dan memprediksi kualitas berdasarkan faktor-faktor yang diidentifikasi melalui clustering. Kombinasi kedua metodologi ini memungkinkan penelitian untuk memanfaatkan kelebihan masing-masing dalam analisis data yang kompleks dan berdimensi tinggi, seperti yang sering ditemukan dalam data akademik perguruan tinggi.

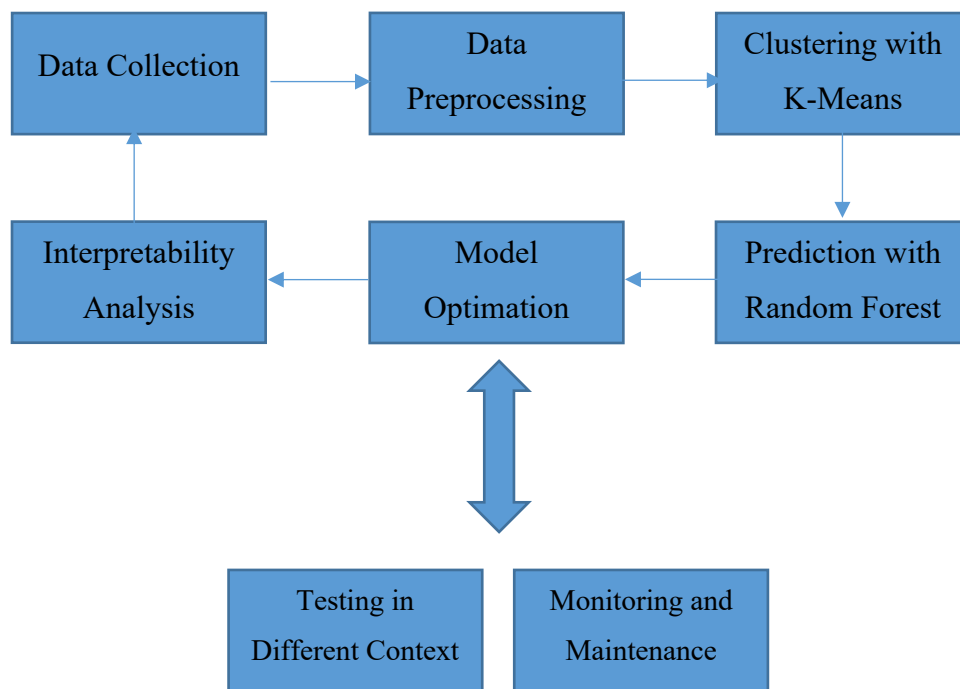
3 METODE PENELITIAN

3.1 Motivasi

Motivasi utama dari penelitian ini adalah untuk meningkatkan kualitas pendidikan tinggi di Indonesia melalui penerapan teknologi data science yang lebih akurat dan dapat diandalkan. Dalam era digital, data yang melimpah dari berbagai sumber, seperti PDDikti, memerlukan analisis yang cerdas untuk menghasilkan wawasan yang bermakna bagi pengambilan keputusan. Model hybrid yang menggabungkan K-Means dan Random Forest diharapkan mampu mengatasi keterbatasan metode sebelumnya dengan memberikan hasil prediksi yang lebih baik dan interpretabilitas yang lebih tinggi. Selain itu, penelitian ini bertujuan untuk memberikan alat yang berguna bagi Dirjen Pendidikan Tinggi untuk memantau, mengevaluasi, dan meningkatkan sistem penjaminan mutu di perguruan tinggi, sehingga mampu menciptakan lingkungan pendidikan yang lebih transparan, efisien, dan responsif terhadap kebutuhan pemangku kepentingan. Kebutuhan akan pendekatan yang lebih inovatif dan berbasis data dalam penjaminan mutu pendidikan tinggi menjadi dorongan kuat untuk melakukan penelitian ini, memastikan bahwa kebijakan yang diambil benar-benar didasarkan pada analisis yang komprehensif dan akurat.

3.2 Framework Riset

Penelitian dilakukan dalam beberapa tahapan. Mulai dari tahapan pencarian literatur hingga pemodelan dan perhitungan. Hasil pengujian diharapkan dapat memenuhi persyaratan tertentu guna menjawab tujuan dari penelitian ini. Tahapan metode penelitian dalam bentuk diagram alir dapat dilihat pada Gambar 2.



Gambar 2 Diagram Alir Metode Penelitian Citra Digital

1. Pengumpulan Data (Data Collection)

Data dikumpulkan dari PDDikti yang mencakup indikator kelembagaan dan aktivitas akademik perguruan tinggi. Data eksternal seperti peringkat internasional dan data alumni juga dapat digunakan jika tersedia. Format data dalam bentuk database relasional yang dapat diubah dalam bentuk CSV sesuai kebutuhan pemrosesan.

2. Pra-Pemrosesan Data (Data Preprocessing)

Pembersihan Data: Hapus nilai yang hilang, duplikasi, dan data yang tidak konsisten untuk meningkatkan kualitas data.

Normalisasi Data: Skalikan fitur untuk memastikan mereka berada dalam rentang yang sama, menggunakan teknik seperti min-max scaling atau z-score normalization.

Transformasi Data: Lakukan transformasi yang diperlukan, seperti encoding variabel kategoris menggunakan One-Hot Encoding atau Label Encoding.

3. Klasterisasi dengan K-Means

Penentuan Jumlah Klaster: Gunakan metode Elbow Method atau Silhouette Score untuk menentukan jumlah klaster yang optimal.

Implementasi K-Means: Terapkan algoritma K-Means untuk mengelompokkan data perguruan tinggi berdasarkan indikator yang telah dikumpulkan.

Evaluasi Klasterisasi: Evaluasi hasil klasterisasi dengan metrik seperti Silhouette Score atau Davies-Bouldin Index untuk memastikan kualitas klaster.

4. Prediksi dengan Random Forest (Prediction with Random Forest)

Pembuatan Fitur: Gunakan hasil klasterisasi dari K-Means sebagai fitur tambahan dalam model Random Forest.

Pemilihan Fitur: Identifikasi fitur-fitur penting yang akan digunakan dalam model prediksi, seperti indikator tata pamong, aktivitas akademik, dan klaster hasil K-Means.

Pelatihan Model: Latih model Random Forest menggunakan data pelatihan. Gunakan teknik cross-validation untuk menghindari overfitting.

Evaluasi Model: Evaluasi kinerja model menggunakan metrik seperti akurasi, precision, recall, dan F1-Score.

5. Optimasi Model (Model Optimization)

Tuning Parameter: Lakukan tuning parameter untuk K-Means dan Random Forest menggunakan Grid Search atau Random Search untuk mendapatkan parameter optimal.

Validasi Model: Validasi model yang telah dioptimasi menggunakan data uji terpisah untuk memastikan generalisasi model.

6. Analisis Interpretabilitas (Interpretability Analysis)

Teknik Interpretabilitas: Gunakan teknik seperti SHAP (Shapley Additive Explanations) atau LIME (Local Interpretable Model-agnostic Explanations) untuk meningkatkan interpretabilitas model Random Forest.

Visualisasi Hasil: Visualisasikan hasil klasterisasi dan prediksi untuk memudahkan pemahaman oleh pemangku kepentingan.

7. Uji Coba di Konteks Berbeda (Testing in Different Contexts)

Evaluasi Geografis dan Temporal: Uji model di berbagai konteks geografis dan temporal untuk mengevaluasi robustnes dan generalisasinya.

Penyesuaian Model: Lakukan penyesuaian pada model berdasarkan hasil evaluasi untuk memastikan model dapat diterapkan secara luas.

8. Implementasi dan Penggunaan (Implementation and Usage)

Integrasi Sistem: Integrasikan model yang telah dikembangkan ke dalam sistem penjaminan mutu perguruan tinggi untuk digunakan oleh Dirjen Pendidikan Tinggi dan pemangku kepentingan lainnya.

Pelatihan Pengguna: Berikan pelatihan kepada pengguna untuk memastikan mereka dapat menggunakan dan memahami hasil dari model tersebut.

9. Monitoring dan Pemeliharaan (Monitoring and Maintenance)

Pemantauan Kinerja: Pantau kinerja model secara berkala untuk memastikan model tetap akurat dan relevan.

Pembaruan Model: Lakukan pembaruan pada model berdasarkan data baru atau perubahan dalam indikator penjaminan mutu untuk memastikan model tetap up-to-date.

3.3 Pendekatan

Pendekatan yang dilakukan dalam penelitian ini meliputi:

1. Optimasi Hasil Klasterisasi dan Prediksi:

Pendekatan ini memastikan bahwa baik klasterisasi maupun prediksi dilakukan dengan akurasi tinggi, memberikan hasil yang dapat diandalkan untuk pengambilan keputusan.

2. Integrasi Data Eksternal:

Dengan menggunakan data eksternal, model menjadi lebih kaya dan memberikan wawasan yang lebih mendalam, yang sangat penting untuk penjaminan mutu pendidikan tinggi.

3. Interpretabilitas Model:

Menggunakan teknik interpretabilitas seperti SHAP atau LIME memastikan bahwa hasil model dapat dipahami oleh pemangku kepentingan, meningkatkan kepercayaan dan adopsi model.

DAFTAR PUSTAKA

Beaulac, C. and Rosenthal, J.S. (2019) ‘Predicting University Students’ Academic Success and Major Using Random Forests’, *Research in Higher Education* [Preprint].

Bicego, M. (2019) ‘K-Random Forests: a K-means style algorithm for Random Forest clustering’, *IEEE International Joint Conference on Neural Network* [Preprint].

Dai, J. (2022) ‘Improving Random Forest Algorithm for University Academic Affairs Management System Platform Construction’, *Advances in Multimedia* [Preprint].

Deepika, K. and Sathyanarayana, N. (2019) ‘Relief-F and Budget Tree Random Forest Based Feature Selection for Student Academic Performance Prediction’, *International Journal of Intelligent Engineering and Systems* [Preprint].

Ghosh, S. and Janan, F. (2021) ‘Prediction of Student’s Performance Using Random Forest Classifier’, *Proceedings of the International Conference on Industrial Engineering and Operations Management* [Preprint].

Liang, J. *et al.* (2019) ‘An Improved Algorithm based on KNN and Random Forest’, *International Conference on Computer Science and Application Engineering* [Preprint].

Prasetyowati, M.I., Maulidevi, N. and Surendro, K. (2020) ‘Feature selection to increase the random forest method performance on high dimensional data’.

Ramadhan, A., Susetyo, B. and Indahwati (2021) ‘Classification Modelling of Random Forest to Identify the Important Factors in Improving the Quality of Education’, *International Journal on Advanced Science, Engineering and Information Technology* [Preprint].

Wiyono, S. *et al.* (2019) ‘Feature Selection With the Random Forest Packages to Predict Student Performance’, *Systemic Information System and Informatics Journal* [Preprint].

Yi, H. *et al.* (2019) ‘A Novel Random Forest and its Application on Classification of Air Quality’, *IIAI International Conference on Advanced Applied Informatics* [Preprint].

Zhang, C. *et al.* (2021) ‘A Modified Random Forest Based on Kappa Measure and Binary Artificial Bee Colony Algorithm’, *IEEE Access* [Preprint].

Zhang, H. and Bu, F. (2019) ‘Weighted KNN Algorithm Based on Random Forests’, *International Conference on Machine Learning and Computing* [Preprint].