

RISIKO KOLABORASI R&D DENGAN PENDEKATAN ALGORITMA C4.5 DAN RANDOM FOREST

Roni Habibi¹, Darfial Guslan², Rd. Nuraini Siti Fatonah³
D4 Teknik Informatika, Universitas Logistik dan Bisnis Internasional¹
D4 Logistik Bisnis, Universitas Logistik dan Bisnis Internasional²
D4 Teknik Informatika, Universitas Logistik dan Bisnis Internasional³

roni.habibi@ulbi.ac.id¹, darfial@ulbi.ac.id², nuraini@ulbi.ac.id³

Abstrak

Studi ini menyelidiki penerapan algoritma C4.5 dan Random Forest dalam mengelola risiko dalam konteks kolaborasi R&D. Mengingat semakin kompleksnya lanskap global, yang ditandai dengan beragamnya budaya, tujuan, dan tantangan berbagi pengetahuan, manajemen risiko yang efektif dalam kolaborasi penelitian dan pengembangan sangatlah penting. Pendekatan kuantitatif, menggunakan metodologi CRISP-DM (Proses Standar Lintas Industri untuk Penambangan Data), diadopsi untuk menyusun proses pengumpulan dan analisis data. Kuesioner khusus digunakan untuk mengumpulkan data, mengidentifikasi risiko kolaborasi R&D dan faktor-faktor yang mempengaruhi kolaborasi ini. Selain itu, penelitian ini mengevaluasi kemanjuran algoritma C4.5 dan Random Forest dalam mengatasi risiko-risiko ini. Data yang dikumpulkan dikenakan analisis statistik, termasuk analisis deskriptif dan klasifikasi. Pendekatan kuantitatif ini bertujuan untuk memberikan wawasan yang obyektif dan terukur mengenai penerapan praktis algoritma C4.5 dan Random Forest untuk manajemen risiko dalam kolaborasi penelitian dan pengembangan. Hasil penelitian menunjukkan bahwa algoritma C4.5 dan Random Forest menunjukkan potensi yang kuat dalam memitigasi risiko yang terkait dengan kolaborasi penelitian dan pengembangan secara efektif. Temuan ini diharapkan dapat memberikan panduan praktis bagi organisasi-organisasi yang terlibat dalam kolaborasi penelitian dan pengembangan, membekali mereka dengan strategi manajemen risiko yang lebih baik dan pendekatan yang lebih efektif untuk mengatasi tantangan-tantangan yang muncul.

Kata kunci: C4.5, Kolaborasi R&D, Random Forest, Risiko.

Abstract

This study investigates the application of the C4.5 and Random Forest algorithms in managing risks within the context of R&D collaboration. In light of the growing complexity of the global landscape, characterized by diverse cultures, objectives, and knowledge-sharing challenges, effective risk management in R&D collaboration is of paramount importance. A quantitative approach, employing the CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology, is adopted to structure the data collection and analysis process. A dedicated questionnaire was employed to collect data, identifying R&D collaboration risks and factors influencing these collaborations. Additionally, this study evaluates the efficacy of the C4.5 and Random Forest algorithms in addressing these risks. The collected data is subjected to statistical analysis, including both descriptive and classification analyses. This quantitative approach aims to provide objective and quantifiable insights into the practical application of the C4.5 and Random Forest algorithms for risk management in R&D collaboration. The results of the research demonstrate that both the C4.5 and Random Forest algorithms exhibit strong potential in effectively mitigating risks associated with R&D collaboration. These findings are expected to offer practical guidance to organizations engaged in R&D collaboration, equipping them with enhanced risk management strategies and more effective approaches to address the challenges that arise.

Keywords: C4.5, Collaboration R&D, Random Forest, Risk

I. PENDAHULUAN

Dalam era globalisasi yang semakin maju, kolaborasi penelitian dan pengembangan (R&D) telah menjadi tonggak penting bagi organisasi yang menginginkan menghasilkan inovasi unggul dan mencapai keunggulan kompetitif. Melalui kolaborasi R&D, berbagai pihak yang terlibat, termasuk lembaga penelitian, universitas, industri, dan pemerintah, dapat bersatu untuk mengatasi masalah yang kompleks dan menciptakan solusi yang baru dan canggih. Namun, seperti halnya kerja sama lintas batas, kolaborasi R&D juga membawa risiko yang signifikan yang harus dikelola secara efektif.

Risiko-risiko ini mencakup perbedaan budaya di antara pihak-pihak yang terlibat, yang dapat mengakibatkan kesulitan dalam pemahaman dan komunikasi. Selain itu, terdapat perbedaan tujuan dan kepentingan antara mitra kolaborasi yang mungkin menghambat keselarasan dalam mencapai tujuan bersama. Kesulitan dalam berbagi pengetahuan dan mengintegrasikan sumber daya serta keahlian yang beragam juga merupakan tantangan yang sering ditemui dalam kolaborasi R&D. Maka, untuk mencapai keberhasilan dan meminimalkan risiko-risiko ini, diperlukan pendekatan manajemen risiko yang kuat dan efektif.

Penelitian ini berfokus pada pengembangan pendekatan manajemen risiko dalam kolaborasi R&D, dan memanfaatkan algoritma C4.5 dan Random Forest. Data survei yang bersifat rule-based memiliki struktur yang terorganisir berdasarkan aturan-aturan tertentu. Hal ini membuatnya sangat sesuai untuk dianalisis menggunakan metode pohon keputusan seperti C4.5 dan algoritma Random Forest. Pada awalnya, survei ini mungkin telah dirancang dengan aturan-aturan yang jelas untuk memandu proses pengumpulan data. Aturan-aturan ini mungkin mencakup bagaimana pertanyaan diajukan, bagaimana respon diklasifikasikan, dan bagaimana data direkam.

Algoritma C4.5, yang dikenal luas dalam bidang data mining, digunakan untuk mengidentifikasi dan mengklasifikasikan berbagai faktor risiko yang mungkin muncul dalam konteks kolaborasi R&D. Algoritma ini memungkinkan identifikasi perbedaan budaya, perbedaan tujuan, dan perbedaan kepentingan yang mungkin menjadi pemicu risiko. Sebagai contoh, dapat dengan efektif mengurai aturan-aturan ini dan membantu mengidentifikasi pola-pola yang bermanfaat. Pohon keputusan C4.5 akan memecah data berdasarkan aturan-aturan ini, memungkinkan analisis yang lebih dalam terkait faktor-faktor apa yang memiliki dampak signifikan pada hasil survei. Dengan cara ini, metode ini membantu menghasilkan wawasan yang lebih kaya dan mendalam mengenai variabel-variabel yang mempengaruhi hasil survei. Sementara itu, algoritma Random Forest, sebagai metode ensemble learning, digunakan untuk mengembangkan model prediksi risiko yang berdasarkan berbagai atribut yang relevan. Ini mencakup karakteristik kolaborasi, faktor

lingkungan yang dapat mempengaruhi kelangsungan kolaborasi, dan kepentingan pihak-pihak yang terlibat. Algoritma ini memungkinkan organisasi untuk mengantisipasi dan mengurangi risiko dengan lebih akurat. Pendekatan ini sangat bermanfaat untuk mengatasi kompleksitas dan kebisingan dalam data survei, memastikan bahwa model yang dihasilkan lebih stabil dan mampu beradaptasi dengan baik terhadap berbagai situasi.

Dengan menggunakan pendekatan algoritma C4.5 dan Random Forest, penelitian ini bertujuan untuk menciptakan sebuah framework manajemen risiko yang dapat membantu organisasi dalam mengidentifikasi, menganalisis, dan mengurangi risiko-risiko yang mungkin muncul selama proses kolaborasi R&D. Framework ini akan membantu meningkatkan efektivitas dan keberhasilan kolaborasi tersebut. Dengan seperti ini maka hasil lebih maksimal dalam analisis data survei yang bersifat rule-based, membuka potensi untuk pengambilan keputusan yang lebih tepat dan efektif. Dengan memahami aturan-aturan yang mendasari data survei, dapat memperoleh wawasan yang lebih mendalam tentang preferensi, perilaku, atau opini responden. Ini adalah langkah penting dalam membimbing tindakan lanjut dan pengambilan keputusan yang terinformasi dalam konteks survei tersebut.

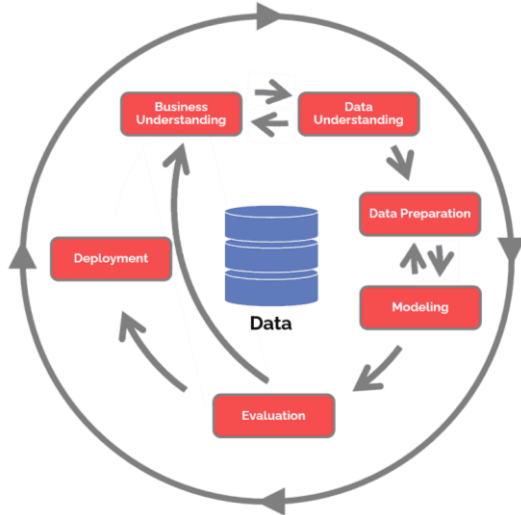
II. METODE

Metode penelitian yang digunakan dalam penelitian ini adalah pendekatan kuantitatif dengan menggunakan kuesioner. Penelitian ini difokuskan pada mahasiswa di lingkungan universitas. Selain itu, penelitian ini mengintegrasikan metode R&D collaboration dan CRISP-DM (Cross-Industry Standard Process for Data Mining) untuk memperdalam pemahaman tentang manajemen risiko kolaborasi R&D di lingkungan universitas.

Metode R&D collaboration digunakan dalam penelitian ini untuk merumuskan pertanyaan penelitian yang spesifik dan relevan terkait dengan manajemen risiko kolaborasi R&D di lingkungan universitas. Metode ini melibatkan kerjasama antara peneliti, mahasiswa, dan staf universitas. Melalui kolaborasi ini, peneliti dapat memperoleh wawasan dan pemahaman yang lebih baik tentang topik penelitian. Dengan melibatkan mahasiswa, penelitian ini juga dapat memberikan peluang bagi mereka untuk berpartisipasi dalam proses penelitian dan memperdalam pemahaman mereka tentang manajemen risiko kolaborasi R&D.

Selanjutnya, dilakukan perancangan kuesioner yang mencakup pertanyaan-pertanyaan terkait dengan variabel yang ingin diteliti. Pertanyaan-pertanyaan ini meliputi risiko-risiko dalam kolaborasi R&D, faktor-faktor yang mempengaruhi keberhasilan kolaborasi, penggunaan pendekatan algoritma C4.5 dan Random Forest dalam analisis risiko, serta persepsi tentang efektivitas framework manajemen risiko yang diusulkan.

Metode CRISP-DM (Cross-Industry Standard Process for Data Mining) digunakan dalam tahap analisis data. Kerangka kerja CRISP-DM terdiri dari enam tahap utama yang membantu dalam mengorganisasi dan memahami proses pengumpulan dan analisis data yang terkait dengan manajemen risiko kolaborasi R&D di lingkungan universitas.



Gambar 1. Metodologi CRIP-DM

1. *Business Understanding* : Pada tahap ini, peneliti memperoleh pemahaman yang mendalam tentang konteks bisnis dalam manajemen risiko kolaborasi R&D di lingkungan universitas. Tujuan penelitian, kebutuhan informasi, dan tantangan yang dihadapi dalam manajemen risiko kolaborasi R&D dipelajari dengan cermat. Hal ini membantu peneliti dalam merumuskan pertanyaan penelitian yang relevan dan menentukan arah penelitian.

2. *Data Understanding* : Pada tahap ini, peneliti memperoleh pemahaman tentang data yang diperlukan dalam penelitian ini. Peneliti menganalisis jenis data yang diperlukan, sumber data yang tersedia, dan potensi data yang relevan untuk memahami manajemen risiko kolaborasi R&D di lingkungan universitas.

3. *Data Preparation* : Tahap ini melibatkan pemrosesan dan pengolahan data yang diperoleh melalui kuesioner. Peneliti membersihkan, mengintegrasikan, dan mempersiapkan data agar siap digunakan dalam tahap analisis lebih lanjut. Langkah-langkah ini termasuk dalam tahap eksplorasi dan persiapan data untuk memastikan kualitas dan konsistensi data.

4. *Modelling* : Pada tahap ini, peneliti menggunakan pendekatan algoritma C4.5 dan Random Forest untuk menganalisis risiko kolaborasi R&D di lingkungan universitas. Pendekatan ini membantu dalam pemodelan dan identifikasi risiko yang mungkin muncul serta membantu dalam pengambilan keputusan yang berhubungan dengan manajemen risiko.

5. *Evaluation* : Tahap ini melibatkan penilaian terhadap hasil analisis yang telah dilakukan. Peneliti mengevaluasi hasil analisis risiko, termasuk keakuratan dan kredibilitas

model yang digunakan, serta mengukur efektivitas framework manajemen risiko yang diusulkan dalam lingkungan universitas. Evaluasi ini penting untuk memastikan kesesuaian dan keandalan hasil penelitian.

6. *Deployment* : Pada tahap ini, peneliti menyusun laporan penelitian yang mencakup temuan dan hasil dari penelitian ini. Selain itu, peneliti juga menyampaikan temuan kepada pihak-pihak terkait, seperti mahasiswa dan staf universitas, untuk memberikan kontribusi dalam pengembangan bidang manajemen risiko kolaborasi R&D di lingkungan universitas. Dengan cara ini, penelitian ini berfungsi sebagai sumber pengetahuan yang dapat digunakan dalam konteks praktis dan akademis.

Penelitian ini berlokasi di Universitas Logistik Bisnis Internasional yang menjadi lingkungan penelitian yang relevan dengan kolaborasi R&D. Dalam penelitian ini, variabel yang diamati dan diukur mencakup persepsi, pengalaman, sikap, dan pendapat mahasiswa terkait manajemen risiko kolaborasi R&D.

Melalui metode R&D collaboration dan penggunaan kerangka kerja CRISP-DM, diharapkan penelitian ini dapat menghasilkan temuan yang lebih komprehensif dan mendalam tentang manajemen risiko kolaborasi R&D di lingkungan universitas, serta memberikan kontribusi penting dalam pengembangan bidang ini.

A. Manajemen

Manajemen adalah proses perencanaan, pengorganisasian, pengarahan, pengendalian, dan pengkoordinasian sumber daya dalam suatu organisasi atau entitas dengan tujuan mencapai hasil yang diinginkan[1]. Manajemen melibatkan penggunaan berbagai teknik, metode, dan prinsip untuk mengoptimalkan penggunaan sumber daya, mengarahkan kegiatan, memimpin orang-orang, mengambil keputusan, dan mencapai tujuan yang telah ditetapkan.

Tujuan utama dari manajemen adalah mencapai efisiensi dan efektivitas[2]. Efisiensi berarti melakukan kegiatan dengan menggunakan sumber daya yang ada secara optimal, sedangkan efektivitas berarti mencapai hasil yang diinginkan sesuai dengan tujuan yang telah ditetapkan.

B. Manajemen Risiko

Manajemen risiko adalah proses identifikasi, penilaian, pengendalian, dan pemantauan risiko dalam suatu organisasi atau entitas dengan tujuan mengurangi atau mengelola dampak negatif dari risiko dan memanfaatkan peluang yang muncul[3]. Manajemen risiko melibatkan penggunaan metodologi dan pendekatan yang sistematis untuk mengidentifikasi risiko potensial, menganalisis dampaknya, mengevaluasi kemungkinan terjadinya, dan mengambil langkah-langkah yang tepat untuk mengurangi risiko atau memanfaatkan peluang yang ada.

Pada dasarnya, manajemen risiko bertujuan untuk melindungi nilai organisasi dengan mengurangi kerugian yang mungkin terjadi akibat risiko dan memaksimalkan keuntungan yang bisa didapatkan dari peluang yang ada[4].

C. Kolaborasi R&D

Kolaborasi R&D (Research and Development) adalah suatu bentuk kerjasama yang dilakukan antara dua atau lebih entitas, baik itu organisasi, lembaga, atau individu, untuk melakukan kegiatan riset dan pengembangan bersama[5]. Kolaborasi R&D melibatkan berbagai pihak yang saling berbagi pengetahuan, sumber daya, dan keahlian dalam rangka menciptakan inovasi, mengembangkan produk atau layanan baru, atau mengeksplorasi pengetahuan baru dalam suatu bidang.

Kolaborasi R&D dapat dilakukan dalam berbagai bentuk, seperti konsorsium riset, kemitraan industri-akademik, proyek kolaboratif, atau jaringan riset[6]. Dalam konteks penelitian ini, kolaborasi R&D di Universitas Logistik Bisnis Internasional bertujuan untuk mengelola risiko dalam kegiatan kolaboratif yang melibatkan berbagai pihak terkait, seperti mahasiswa, dosen, dan stakeholder lainnya.

D. Komunikasi

Komunikasi adalah proses penyampaian dan pertukaran informasi, gagasan, pikiran, dan perasaan antara dua atau lebih pihak[7]. Tujuan utama komunikasi adalah untuk memahami dan dipahami oleh pihak-pihak yang terlibat. Komunikasi melibatkan penggunaan berbagai saluran, metode, dan simbol untuk menyampaikan pesan secara efektif.

Dalam konteks manajemen, komunikasi memiliki peran yang sangat penting[8]. Komunikasi yang efektif memungkinkan para manajer dan karyawan untuk berinteraksi, berbagi informasi, dan bekerja sama dalam mencapai tujuan organisasi. Komunikasi yang baik juga membantu membangun hubungan yang sehat antara individu-individu dalam organisasi.

E. Algoritma C4.5

Algoritma C4.5 adalah algoritma pembelajaran mesin yang digunakan dalam analisis data dan pengambilan keputusan[9]. Algoritma ini digunakan untuk menghasilkan model pohon keputusan berdasarkan data historis yang tersedia. Model pohon keputusan ini dapat digunakan untuk mengklasifikasikan data baru berdasarkan aturan yang ditemukan dari data pelatihan.

Algoritma C4.5 menggunakan metode rekursif untuk membangun pohon keputusan. Pada setiap tahap, algoritma C4.5 memilih atribut yang memiliki kemampuan paling tinggi untuk memisahkan data menjadi kelompok yang homogen. Atribut tersebut akan menjadi node pada pohon keputusan. Algoritma kemudian melakukan pemisahan berdasarkan nilai-nilai atribut tersebut dan terus menerus membangun cabang-cabang pohon keputusan hingga mencapai kondisi berhenti (misalnya, semua data pada cabang tertentu sudah memiliki kelas yang sama atau sudah mencapai kedalaman maksimum).

F. Random Forest

Random Forest adalah sebuah metode ensemble learning yang digunakan dalam analisis data dan

pengambilan keputusan. Metode ini memanfaatkan kombinasi dari beberapa pohon keputusan (decision tree) yang bekerja secara independen untuk menghasilkan prediksi yang lebih akurat[10].

Proses pembentukan Random Forest dimulai dengan pembentukan beberapa pohon keputusan yang dibangun berdasarkan sampel acak dari data pelatihan. Setiap pohon keputusan dihasilkan melalui proses pembangunan yang mirip dengan algoritma C4.5, yaitu pemilihan atribut terbaik untuk memisahkan data pada setiap simpul (node) pohon. Namun, perbedaannya terletak pada penggunaan sampel acak dari data pelatihan, sehingga setiap pohon memiliki variasi dan independensi dalam pembentukannya.

III. HASIL DAN PEMBAHASAN

A. Deskripsi Data

Data yang kami gunakan dalam penelitian ini memiliki beberapa atribut yang memberikan informasi rinci tentang mahasiswa dan alumni ULBI serta pengalaman mereka selama program magang. Beberapa atribut utama dalam dataset ini adalah:

TABEL I
ATRIBUT DAN ARTINYA DARI DATASET

No.	Atribut	Deskripsi
1.	Nama Mahasiswa	Kolom ini berisi nama lengkap dari mahasiswa atau alumni ULBI yang telah mengikuti program magang.
2.	Jenis Kelamin	Kolom ini berisi nama lengkap dari mahasiswa atau alumni ULBI yang telah mengikuti program magang.
3.	Angkatan	Tahun masuk mahasiswa ke ULBI. Ini memberikan informasi tentang generasi mahasiswa.
4.	Umur	Usia mahasiswa atau alumni saat mereka menjalani program magang.
5.	Program Studi- Jurusan	Menunjukkan program studi atau jurusan yang diambil oleh mahasiswa atau alumni, seperti "D4-Teknik Informatika," "D3-Manajemen Logistik," dan berbagai program lainnya.
6.	Perusahaan Tempat Magang	Nama perusahaan atau lembaga tempat mahasiswa atau alumni melakukan program magang. Informasi ini membantu mengidentifikasi perusahaan atau organisasi yang terlibat dalam kolaborasi dengan ULBI

7.	Bidang Usaha Perusahaan	Mendesripsikan jenis bisnis atau bidang usaha dari perusahaan tempat mahasiswa atau alumni melakukan program magang, seperti "Logistik," "Pengembang & Operator Real Estat," atau bidang usaha lainnya.
8.	Jangka Waktu Magang (per)	Kolom ini mencatat berapa lama mahasiswa atau alumni menjalani program magang dalam satuan bulan.
9.	Kesesuaian Minat dan Bakat	Penilaian tingkat kesesuaian program magang dengan minat dan bakat mahasiswa/alumni dalam skala 1-5.
10.	Mudah Memahami Tugas dan Tanggung Jawab	Mengukur kemudahan pemahaman mahasiswa/alumni terhadap tugas dan tanggung jawab yang diberikan selama magang dalam skala 1-5.
11.	Kesesuaian Tugas dan Tanggung Jawab	Menilai tingkat kesesuaian tugas dan tanggung jawab dengan program studi/jurusan mahasiswa/alumni dalam skala 1-5.
12.	Komunikasi dengan Rekan Kerja dan Atasan	Penilaian kemampuan mahasiswa/alumni dalam berkomunikasi dengan rekan kerja dan atasan selama program magang dalam skala 1-5.
13.	Informasi Mengenai Peraturan dan Kebijakan Tempat Kerja	Mengukur pemahaman mahasiswa/alumni terhadap peraturan dan kebijakan tempat magang dalam skala 1-5.
14.	Program Magang Tepat Waktu	Mengindikasikan apakah mahasiswa/alumni menyelesaikan program magang sesuai jadwal yang ditentukan dalam skala 1-5.
15.	Keadilan dalam Evaluasi Kinerja	Penilaian apakah evaluasi kinerja selama program magang dianggap adil oleh mahasiswa/alumni dalam skala 1-5.
16.	Kesesuaian dalam Evaluasi Program Kerja	Menilai sejauh mana program kerja yang diberikan sesuai dengan program studi/jurusan mahasiswa/alumni dalam skala 1-5.
17.	Adanya Diskriminasi	Penilaian apakah mahasiswa/alumni mengalami

	atau Pelecehan	diskriminasi atau pelecehan selama program magang.
18.	Kenyamanan Lingkungan Kerja	Penilaian tentang sejauh mana mahasiswa/alumni merasa nyaman dengan lingkungan kerja selama program magang dalam skala 1-5.
19.	Total	Total akumulasi penilaian dari seluruh aspek di atas.

B. Pengolahan Data

Data yang digunakan dalam penelitian ini telah melalui serangkaian tahap pengolahan untuk memastikan kualitas data yang baik dalam pemodelan machine learning. Berikut adalah tahap-tahap pengolahan data yang telah dilakukan :

- 1) *Penghapusan Data Tidak Lengkap*: Data yang mengandung nilai kosong (missing values) atau data yang tidak lengkap dihapus dari dataset. Hal ini dilakukan untuk memastikan bahwa data yang digunakan dalam analisis adalah data yang lengkap dan dapat dipercaya.
- 2) *Konversi Kolom Teks Menjadi Data Numerik*: Beberapa kolom dalam dataset, seperti "Program Studi-Jurusan," "Bidang Usaha Perusahaan," dan "Jenis Kelamin," awalnya berisi data teks. Untuk digunakan dalam model machine learning, kolom-kolom ini diubah menjadi data numerik. Sebagai contoh, "Program Studi-Jurusan" dikonversi menjadi bilangan bulat, di mana setiap program studi diberi nomor yang sesuai.
- 3) *Konversi Kolom Target*: Kolom "risk_level," yang merupakan target dari model machine learning, berisi tiga kategori tingkat risiko, yaitu "Resiko Rendah," "Resiko Sedang," dan "Resiko Tinggi." Kategori ini diubah menjadi data numerik dengan memberikan label numerik, seperti '1' untuk "Resiko Rendah," '2' untuk "Resiko Sedang," dan '3' untuk "Resiko Tinggi".
- 4) *One-Hot Encoding dan Label Encoding*: Ini adalah teknik yang digunakan untuk mengatasi atribut kategorikal (atribut yang memiliki nilai kategori atau label) sebelum menyematkannya ke dalam model pembelajaran mesin. One-Hot Encoding mengubah atribut kategorikal menjadi vektor biner yang dapat digunakan dalam model. Label Encoding, di sisi lain, mengubah atribut kategorikal menjadi bilangan bulat (integer). Teknik ini diperlukan karena banyak algoritme pembelajaran mesin hanya dapat bekerja dengan data numerik, dan atribut kategorikal perlu dikonversi sebelum digunakan dalam model.
- 5) *Pembagian Data*: Data dibagi menjadi dua subset, yaitu data latih (train) dan data uji (test). Data latih digunakan untuk melatih model machine learning, sementara data uji digunakan untuk menguji sejauh mana model tersebut mampu membuat prediksi yang akurat.

C. Model Machine Learning

Bagian ini berfokus pada penerapan dua jenis model machine learning yang digunakan dalam penelitian, yaitu Random Forest Classifier dan Decision Tree Classifier (C4.5). Setiap model memiliki keunggulan dan cara kerja yang berbeda. Berikut adalah penjelasan lebih rinci tentang kedua model ini:

1) Random Forest Classifier

Random Forest Classifier adalah algoritma machine learning ensemble yang terdiri dari sejumlah besar pohon keputusan (decision trees) yang dikombinasikan. Setiap pohon keputusan digunakan untuk membuat prediksi, dan hasil akhir dari model adalah hasil dari agregasi prediksi dari seluruh pohon. Algoritma ini memiliki beberapa keunggulan:

- Mengatasi Overfitting: Random Forest memiliki mekanisme untuk mengatasi overfitting, yang membuatnya cocok untuk berbagai dataset.
- Pengukuran Fitur Penting: Model ini dapat memberikan informasi tentang fitur-fitur yang paling penting dalam membuat prediksi, yang dapat digunakan untuk analisis lebih lanjut.

```
from sklearn.metrics import
classification_report, accuracy_score
from sklearn.metrics import accuracy_score,
precision_score, recall_score, f1_score,
classification_report, roc_auc_score

y_pred= RF.predict(x_test)
print(classification_report(y_test, y_pred))

Output :
```

	precision	recall	f1-score	support
0	0.00	0.00	0.00	2
1	0.95	1.00	0.98	143
2	1.00	0.94	0.97	78
accuracy			0.97	223
macro avg	0.65	0.65	0.65	223
weighted avg	0.96	0.97	0.96	223

Keterangan :
 Dalam evaluasi model Random Forest Classifier, digunakan metode predict untuk memperoleh prediksi dari data pengujian, kemudian hasil prediksi dibandingkan dengan label sebenarnya yang terdapat dalam y_test. Dalam pengukuran kinerja model, digunakan beberapa metrik evaluasi, termasuk Accuracy Score yang mengukur sejauh mana model dapat mengklasifikasikan data dengan benar, serta Classification Report yang memberikan informasi rinci tentang precision, recall, dan f1-score untuk setiap kelas. Metrik-metrik ini memberikan gambaran yang komprehensif tentang kemampuan model dalam mengklasifikasikan data.

2) Decision Tree Classifier (C4.5)

Decision Tree Classifier adalah algoritma machine learning yang menghasilkan pohon keputusan untuk membuat prediksi. Setiap simpul dalam pohon mewakili keputusan berdasarkan atribut-atribut data. Keunggulan dari Decision Tree Classifier meliputi:

- Interpretasi yang Mudah: Pohon keputusan dapat dengan mudah diinterpretasikan oleh manusia, sehingga mudah dimengerti.
- Penggunaan dalam Pemilihan Fitur: Decision Tree Classifier dapat digunakan untuk pemilihan fitur, yaitu mengidentifikasi fitur-fitur yang paling relevan dalam membuat prediksi.

```
from sklearn.metrics import
classification_report, accuracy_score

y_pred= DT.predict(x_test)
print(classification_report(y_test, y_pred))

Output :
```

	precision	recall	f1-score	support
0	0.00	0.00	0.00	2
1	0.86	0.88	0.87	143
2	0.77	0.76	0.76	78
accuracy			0.83	223
macro avg	0.54	0.55	0.54	223
weighted avg	0.82	0.83	0.83	223

Keterangan :
 model Decision Tree Classifier yang telah disesuaikan dengan parameter terbaik dari hasil pencarian GridSearchCV digunakan untuk melakukan prediksi pada data uji (x_test). Hasil prediksi tersebut kemudian digunakan untuk menghitung beberapa metrik evaluasi, seperti akurasi, presisi, recall, dan lainnya.

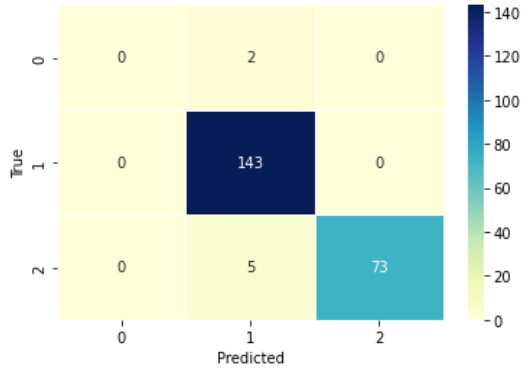
D. Analisis Hasil

Bagian analisis hasil adalah tahap penting dalam penelitian ini karena memberikan wawasan yang berharga tentang bagaimana model-machine learning telah berhasil dalam mengklasifikasikan mahasiswa dan alumni ULBI ke dalam tiga tingkat risiko. Hasil analisis ini juga memberikan gambaran yang lebih jelas tentang dampak kolaborasi R&D.

1) Hasil Evaluasi Model

Random Forest Classifier:

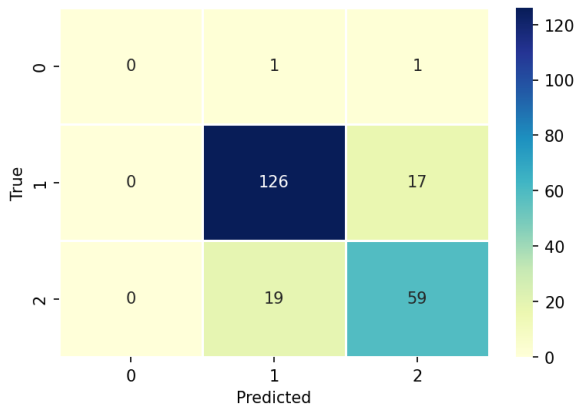
Model ini memberikan tingkat akurasi sekitar 97% pada data uji, yang menggambarkan kemampuannya dalam mengklasifikasikan individu ke dalam tiga kategori risiko dengan tingkat ketepatan yang tinggi. Selain itu, hasil evaluasi menunjukkan bahwa model Random Forest memiliki nilai precision dan recall yang tinggi untuk masing-masing tingkat risiko. Hal ini berarti model ini mampu dengan baik mendeteksi dan mengklasifikasikan individu dengan risiko rendah, sedang, atau tinggi. Hasil ini memberikan indikasi bahwa Random Forest adalah model yang sangat efektif dalam mengelola risiko dalam konteks kolaborasi R&D.



Gambar 2. Confusion Matrix Random Forest

Decision Tree Classifier (C4.5):

Model Decision Tree (C4.5) memberikan akurasi sekitar 83% pada data uji. Meskipun hasil ini masih merupakan tingkat akurasi yang baik, itu adalah lebih rendah dibandingkan dengan Random Forest. Hasil evaluasi juga menunjukkan bahwa model Decision Tree memiliki precision dan recall yang sedikit lebih rendah dibandingkan dengan Random Forest. Namun, tetap penting untuk diingat bahwa model ini masih memberikan kinerja yang baik dalam mengklasifikasikan risiko.



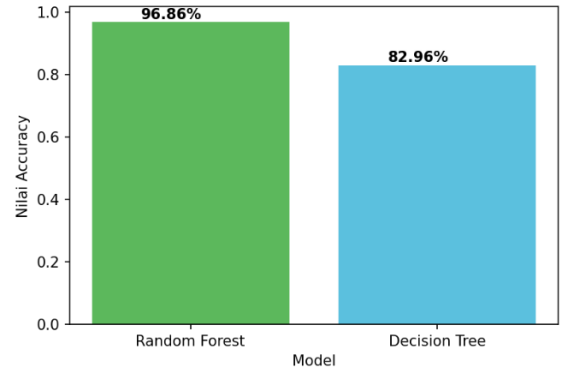
Gambar 3. Confusion Matrix Decision Tree (C4.5)

2) Perbandingan Kinerja Model

Dalam analisis perbandingan kinerja model, hasil evaluasi Random Forest Classifier dan Decision Tree Classifier (C4.5) dibandingkan.

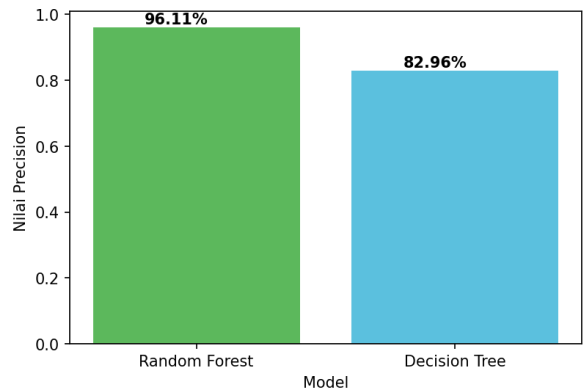
Random Forest vs. Decision Tree:

- **Akurasi:** Random Forest memiliki akurasi yang lebih tinggi, yaitu sekitar 97% dibandingkan dengan 83% dari Decision Tree. Hal ini menunjukkan bahwa Random Forest lebih mampu dalam mengklasifikasikan mahasiswa dan alumni ULBI ke dalam tiga kategori risiko.



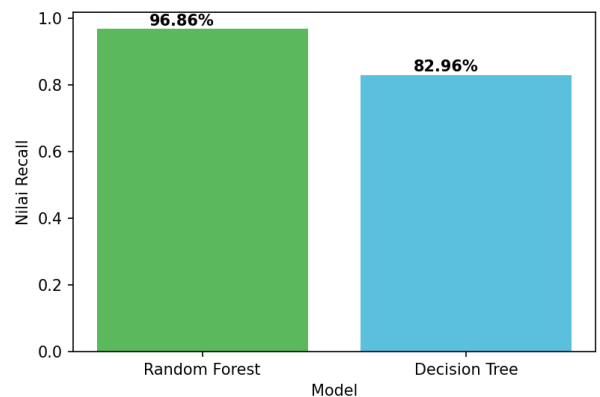
Gambar 4. Perbandingan Nilai Akurasi Random Forest dan Decision Tree (C4.5)

- **Precision:** Precision untuk semua tingkat risiko lebih tinggi pada Random Forest dibandingkan dengan Decision Tree. Ini berarti bahwa ketika Random Forest mengklasifikasikan seseorang ke dalam tingkat risiko tertentu, kemungkinan besar hasil tersebut adalah benar.



Gambar 5. Perbandingan Nilai Precision Random Forest dan Decision Tree (C4.5)

- **Recall:** Recall mengukur kemampuan model untuk mendeteksi individu dengan risiko tertentu. Hasil menunjukkan bahwa Random Forest memiliki recall yang lebih tinggi, yang berarti model ini lebih baik dalam mendeteksi individu berisiko.



Gambar 6. Perbandingan Nilai Recall Random Forest dan Decision Tree (C4.5)

Hasil ini mengindikasikan bahwa Random Forest lebih efektif dalam mengklasifikasikan mahasiswa dan alumni ULBI ke dalam tiga tingkat risiko dibandingkan dengan Decision Tree. Hal ini juga menunjukkan keunggulan model ensemble terhadap model pohon keputusan tunggal dalam konteks penelitian ini.

IV. KESIMPULAN

Penelitian ini bertujuan meningkatkan kolaborasi penelitian di Universitas Logistik dan Bisnis Internasional (ULBI) melalui penggunaan pendekatan C4.5 dan algoritma Random Forest dalam manajemen risiko. Hasil penelitian menunjukkan bahwa algoritma Random Forest lebih efektif daripada Decision Tree (C4.5) dalam mengklasifikasikan tingkat risiko kolaborasi R&D. Implikasinya, ULBI dapat mengidentifikasi mahasiswa dan alumni dengan risiko tinggi atau rendah dalam kolaborasi penelitian untuk menyediakan dukungan yang lebih tepat.

Rekomendasi dari penelitian ini adalah mengimplementasikan algoritma Random Forest dalam solusi pemantauan risiko dan meningkatkan pemahaman tentang algoritma ini melalui pelatihan. Fokus pada dukungan individu dengan risiko tinggi juga penting, dengan bimbingan, sumber daya tambahan, dan program khusus yang mendukung mereka. Penelitian lanjutan dalam pengembangan sistem pemantauan risiko lebih canggih akan membantu ULBI mencapai tujuan kolaboratif yang lebih baik. Hasil penelitian ini memberikan kontribusi penting dalam pengembangan strategi manajemen risiko yang lebih efektif dalam kolaborasi penelitian di universitas, menciptakan inovasi yang bermakna. Kesimpulan dan saran ini membantu ULBI mencapai tujuan kolaboratif mereka di masa depan.

UCAPAN TERIMA KASIH

Terima kasih kepada rekan-rekan di Universitas Logistik dan Bisnis Internasional (ULBI) atas wawasan, data, dan dukungan mereka. Kami juga mengapresiasi partisipasi para responden dalam penelitian ini. Dukungan finansial dari ULBI sangat berarti bagi keberhasilan penelitian ini. Terima kasih juga kepada keluarga, teman-teman, serta para dosen

yang memberikan arahan dan motivasi selama penelitian. Semua kontribusi ini sangat dihargai.

DAFTAR PUSTAKA

- [1] N. Haefner, J. Wincent, V. Parida, and O. Gassmann, "Artificial intelligence and innovation management: A review, framework, and research agenda," *Technological Forecasting and Social Change*, vol. 162, pp. 120392, 2021. ISSN: 0040-1625. doi: 10.1016/j.techfore.2020.120392. [Online]. Available: <https://doi.org/10.1016/j.techfore.2020.120392>.
- [2] K. Zafari, S. Biggemann, and T. Garry, "Mindful management of relationships during periods of crises: A model of trust, doubt and relational adjustments," *Industrial Marketing Management*, vol. 88, pp. 278-286, 2020. ISSN: 0019-8501. doi: 10.1016/j.indmarman.2020.05.026. [Online]. Available: <https://doi.org/10.1016/j.indmarman.2020.05.026>.
- [3] N. V. Syreyshechikova, D. Yu. Pimenov, T. Mikolajczyk, and L. Moldovan, "Development of a Risk Management Technique in Strategic Planning of Universities. Case study of a Polytechnical Institute," in *Procedia Manufacturing*, vol. 46, pp. 256-262, 2020. ISSN: 2351-9789. doi: 10.1016/j.promfg.2020.03.038. [Online]. Available: <https://doi.org/10.1016/j.promfg.2020.03.038>.
- [4] D. Di Ludovico and L. Di Lodovico, "The Regional Management Risk Plan. Knowledge, scenarios and prevention projects in a regional context," *International Journal of Disaster Risk Reduction*, vol. 45, pp. 101465, 2020. ISSN: 2212-4209. doi: 10.1016/j.ijdrr.2019.101465. [Online]. Available: <https://doi.org/10.1016/j.ijdrr.2019.101465>.
- [5] G. Fernandes, J.M.R.C.A. Santos, P. Ribeiro, L.M.D.F. Ferreira, D. O'Sullivan, D. Barroso, and E.B. Pinto, "Critical Success Factors of University-Industry R&D Collaborations," *Procedia Computer Science*, vol. 219, pp. 1650-1659, 2023. ISSN: 1877-0509. doi: 10.1016/j.procs.2023.01.458. [Online]. Available: <https://doi.org/10.1016/j.procs.2023.01.458>.
- [6] S. Kim and K.C. Park, "Government funded R&D collaboration and its impact on SME's business performance," *Journal of Informetrics*, vol. 15, no. 3, pp. 101197, 2021. ISSN: 1751-1577. doi: 10.1016/j.joi.2021.101197. [Online]. Available: <https://doi.org/10.1016/j.joi.2021.101197>.
- [7] Sendjaja, S. D., Rahardjo, T., Pradekso, T., & Sunarwinadi, I. R. (2014). *Teori komunikasi*.
- [8] Effendi, O. U., & Surjaman, T. (1986). *Dinamika komunikasi*. Remadja Karya.
- [9] Mardi, Y. (2017). *Data Mining: Klasifikasi Menggunakan Algoritma C4.5*. *Jurnal Edik Informatika Penelitian Bidang Komputer Sains dan Pendidikan Informatika*, 2(2), 213-219.
- [10] Sihombing, P. R., & Yuliati, I. F. (2021). *Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia*. *MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, 20(2), 417-426.