



Klasifikasi Penentuan Proposal Penelitian Menggunakan Metode Naïve Bayes dan Decision Tree

Saifuddin ¹⁾, EIH Ujianto ²⁾

¹⁾ Jurusan Magister Teknologi Informasi Universitas Teknologi Yogyakarta Sleman, DI Yogyakarta

²⁾ Jurusan Magister Teknologi Informasi Universitas Teknologi Yogyakarta Sleman, DI Yogyakarta

¹⁾Email : saifuddin@student.uty.ac.id

²⁾Email : erik.iman@uty.ac.id

Abstrak – Universitas melalui Lembaga Penelitian dan Pengabdian kepada Masyarakat (LPPM) menyelenggarakan kegiatan penelitian dan pengabdian kepada masyarakat bagi dosen secara profesional dan proporsional. Dalam melakukan review kegiatan wajib menyesuaikan dengan strategi yang telah ditentukan dengan pendampingan LPPM. Pelaksanaan yang dilakukan adalah evaluasi proposal penelitian dan pengabdian kepada masyarakat yang dapat dinyatakan layak atau tidak layak untuk dilaksanakan dan dibiayai dengan dana perguruan tinggi. Setiap gagasan penelitian dan pengabdian kepada masyarakat yang diajukan oleh seorang dosen harus melalui seleksi substansi, yaitu pemilihan proposal penelitian yang diajukan dan diseleksi oleh tim reviewer yang ditunjuk oleh LPPM. Hal ini juga dilakukan oleh LPPM Universitas Tunas Pembangunan (UTP) dalam memutuskan proposal tanpa bantuan mesin. Hasil penelitian yang dilakukan mulai dari tingkat pendahuluan hingga uji coba dengan menggunakan metode tipe Naïve Bayes diperoleh akurasi sebesar 96%, dan pendekatan Choice Tree diperoleh akurasi sebesar 64%.

Kata Kunci : *klasifikasi, penelitian, data mining, dosen, penilaian.*

PENDAHULUAN

Pendidikan tinggi adalah jenjang pendidikan setelah pendidikan menengah yang meliputi program pendidikan diploma, sarjana, magister, spesialis, dan doktoral yang diselenggarakan oleh perguruan tinggi. Setiap perguruan tinggi, melalui Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) menyelenggarakan aktivitas penelitian dan pengabdian kepada masyarakat bagi dosen dengan pengelolaan, penilaian serta pendanaan sepenuhnya yang dilaksanakan secara profesional dan proporsional. LPPM dalam melakukan penelitian dan pengabdian kepada masyarakat wajib sesuai prosedur yang telah ditetapkan salah satunya adalah penilaian proposal penelitian dan pengabdian kepada masyarakat dapat dinyatakan layak atau tidak layak untuk dilaksanakan dan didanai oleh perguruan tinggi. Setiap proposal penelitian dan pengabdian kepada masyarakat yang diajukan oleh dosen wajib melalui tahap seleksi, baik seleksi tahap administrasi maupun seleksi substansi dalam penyeleksian proposal penelitian yang masuk akan diseleksi oleh tim reviewer yang ditetapkan oleh LPPM. Hal ini pun juga dilakukan oleh LPPM Universitas Tunas Pembangunan (UTP) dalam melakukan seleksi proposal tanpa adanya sistem yang membantu. Dalam penelitian kami, algoritma klasifikasi diterapkan pada data dan diminta untuk memperkirakan diagnosis penyakit. Kinerja klasifikasi jaringan saraf tiruan dan pengklasifikasi Naïve Bayes yang diterapkan pada data dengan 9 input dan satu output dihitung dan hasil kinerjanya dibandingkan. Artikel ini menyoroti evaluasi kinerja berdasarkan contoh klasifikasi data yang benar dan salah menggunakan algoritma klasifikasi ANN dan Naïve Bayes (Saritas & Yasar, 2019).

Decision Tree dan Naïve Bayes untuk membandingkan kinerja klasifikasi kumpulan data kami. Kami telah mengambil dataset kinerja siswa yang memiliki 480 pengamatan. Kami telah mengklasifikasikan siswa ini ke dalam kelompok yang berbeda dan kemudian menghitung keakuratan klasifikasi kami dengan menggunakan bahasa R. Decision tree menggunakan metode Divide and Conquer termasuk beberapa aturan yang memudahkan manusia untuk memahaminya (Yadav & Thareja, 2019), Kemudian algoritma klasifikasi Naïve Bayes (NB) dan Decision Tree (DT) digunakan untuk mengklasifikasikan aktivitas manusia seperti: duduk, berdiri, berjalan, duduk turun dan berdiri. Dari hasil, terlihat bahwa akurasi maksimum 89,5% dan 99,9% dicapai dengan menggunakan pengklasifikasi NB dan DT dengan filter



Gaussian. Selain itu, kami juga telah membandingkan hasil yang diperoleh dengan algoritme rekanannya untuk membuktikan keefektifannya (Maswadi et al., 2021).

Evaluasi dilakukan berdasarkan hasil berbasis akurasi, presisi, recall, dan f1-score. Di antara ketiga pengklasifikasi tersebut, kami telah memilih Multinomial Naïve Bayes sebagai model sour best yang kami gunakan mencapai akurasi 91,8% (Rahman & Akter, 2019) kredibilitas calon pelanggan. Hasil penelitian menunjukkan akurasi algoritma Naïve Bayes 77,29%, Decision Tree 89,02% dan Support Vector Machine (SVM) tertinggi 89,86% (Nurajijah & Riana, 2019)

Algoritma yang diusulkan dibandingkan dengan teknik ansambel dan klasifikasi tunggal seperti Random Forest, Bagged NB, Bagged DT, NB, dan DT. Hasil eksperimen atas 52 set data UCI dengan ukuran tas 100 menunjukkan bahwa algoritma yang diusulkan secara signifikan mengungguli algoritma yang ada (Singh & Singh, 2019). Oleh karena itu, dalam penelitian ini analisis sentimen pengguna KRL Commuter Line Jabodetabek pada data media sosial Twitter diambil secara acak sebanyak 127 data sehingga diolah melalui beberapa tahapan, salah satunya untuk menghindari duplikasi data. Pada penelitian ini menggunakan metode Naive Bayes Classifier, KNN dan Decision Tree dengan beberapa tahapan yaitu Convert Emoticon, Cleansing, Case Folding, Tokenizing, Stemming. Hasil pengujian menunjukkan perbedaan antara metode Naive Bayes adalah akurasi 80%, presisi 66,67%, sensitivitas 100%, spesifisitas 66,67% (Romadloni et al., 2019). Terdapat 11 atribut yang digunakan untuk Klasifikasi Data Mining yaitu NPM, Jenis Kelamin, Umur, Jurusan, Kelas, Jabatan, Indeks Prestasi Semester 1, Indeks Prestasi Semester 2, Indeks Prestasi Semester 3, Indeks Prestasi Semester 4 dan Informasi sebagai atribut hasil. Dari hasil evaluasi dan validasi yang telah dilakukan dengan menggunakan tools RapidMiner akurasi metode Decision Tree (C4.5) sebesar 98,04% pada pengujian ke-3. Akurasi Metode Naïve Bayes adalah 96,00% pada tes ke-4. Dan akurasi Metode K-Nearest Neighbor (K-NN) sebesar 90,00% pada pengujian kedua (Budiyantara et al., 2020).

Pada teknik sebelumnya pendekatan pengklasifikasi SVM diterapkan untuk prediksi kinerja siswa. Pendekatan naive bayes diterapkan dalam karya yang diusulkan untuk analisis prediksi kinerja siswa (Tripathi et al., 2019). Hasil perbandingan dari ketiga model tersebut adalah sebagai berikut, K-NN memiliki tingkat akurasi sebesar 77%, Decision Tree memiliki tingkat akurasi sebesar 74%, dan Naive Bayes memiliki tingkat akurasi sebesar 89%. Oleh karena itu, untuk klasifikasi tingkat kelulusan mahasiswa Program Studi Teknik Informatika Universitas Islam Madura direkomendasikan model Naive Bayes karena memiliki tingkat akurasi yang lebih baik dibandingkan K-NN dan Decision Tree (Hozairi et al., 2021).

Analisis dilakukan dengan menggunakan dua algoritma klasifikasi data mining seperti Naive Bayes dan Support Vector Machine. Analisis ini bertujuan untuk memprediksi diabetes menggunakan catatan kesehatan dan membandingkan keakuratan kedua algoritma ini untuk menemukan algoritma yang lebih baik untuk memprediksi diabetes (Sindu & Prakash, 2021).

Algoritma yang digunakan dalam penelitian ini adalah Naive Bayes, K-Nearest Neighbor (KNN) dan Decision Tree. Ketiga algoritma tersebut dibandingkan untuk melihat akurasinya dengan menggunakan software Rapidminer. Berdasarkan akurasi tersebut, didapatkan bahwa algoritma K-NN paling baik dalam memprediksi kelulusan mahasiswa dengan akurasi sebesar 91,82% (Hardiani, 2021).

Machine Learning adalah ilmu (seni) pemrograman komputer, sehingga mereka bisa belajar dari data (Westphal et al., 2019). Pembelajaran mesin merupakan sistem yang dapat belajar dengan sendirinya untuk memutuskan sesuatu tanpa harus berulang kali dilakukan pemrograman oleh manusia, sehingga computer menjadi lebih pintar dan dapat belajar dari data yang dimilikinya. Menurut (Retnoningsih & Pramudita, 2020) Machine learning bekerja apabila tersedia data sebagai input untuk dilakukan analisis terhadap kumpulan data besar sehingga menemukan pola tertentu. Data adalah bahan input yang digunakan untuk melakukan pembelajaran (training) sehingga mesin dapat menghasilkan analisis yang benar. Pada machine learning terdapat data training dan data testing, data training digunakan melatih algoritma didalam machine learning sedangkan data testing digunakan untuk mengetahui performa algoritma dalam machine



learning. Sedangkan unsupervised learning merupakan teknik yang dilakukan dengan menarik kesimpulan berdasarkan input data labeled response (Buslim & Iswara, 2019). Model klasifikasi machine learning dibagi menjadi dua yaitu a) Melatih data lalu membedakannya kedalam class; b) Melakukan evaluasi terhadap objek yang tidak dikenal (Buslim & Iswara, 2019). Klasifikasi merupakan bagian unsupervised learning, yang memerlukan persiapan data yang baik serta pemilihan bahasa pemrograman yang mampu mendukung machine learning. Menurut (Aprilla et al., 2013), kata mining merupakan kiasan dari bahasa Inggris, mine. Mine memiliki arti menambang sumber daya yang tersembunyi di dalam tanah, dimana berarti data mining merupakan penggalian makna yang tersembunyi dari kumpulan data yang sangat besar. Karena itu data mining sebenarnya memiliki akar yang panjang dari bidang ilmu seperti kecerdasan buatan (artificial intelligent), machine learning, statistik dan basis data. Naïve Bayes merupakan metode klasifikasi yang statistik berdasarkan teorema Bayes (Ahmed et al., 2018). Naïve Bayes berpotensi baik untuk mengklasifikasikan data karena kesederhanaannya (Panda et al., 2010). Naive Bayes Classifier merupakan metode klasifikasi yang berdasarkan teorema Bayes. Metode klasifikasi ini menggunakan probabilitas dan statistik yang pertama kali ditemukan oleh ilmuwan bangsawan Inggris bernama Thomas Bayes, yaitu suatu metode untuk memprediksi peluang di masa yang akan datang berdasarkan kejadian di masa sebelumnya, sehingga metode ini dikenal sebagai Teorema Bayes. Ciri utama dari Naive Bayes Classifier ini adalah asumsi yang sangat kuat terhadap kondisi atau kejadian (Mustafa et al., 2018). Keuntungan dari penggunaan metode Naïve Bayes adalah hanya membutuhkan sejumlah kecil data untuk memperkirakan parameter (sarana dan varians dari variabel-variabel) klasifikasi. Karena variabel yang diasumsikannya, hanya varian dari variabel untuk beberapa kelas yang ditentukan (Oktasari et al., 2016). Bayesian classification merupakan pengklasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas isi dari suatu class. Bayesian classification didasarkan pada teorema Bayes yang memiliki kemampuan klasifikasi serupa dengan neural network dan decision tree. Bayesian classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam database dengan data yang besar (Jananto, 2013).

$$P(H | X) = \frac{P(X|H)P(H)}{P(X)}$$

Naïve Bayes memiliki persamaan seperti berikut ini :

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

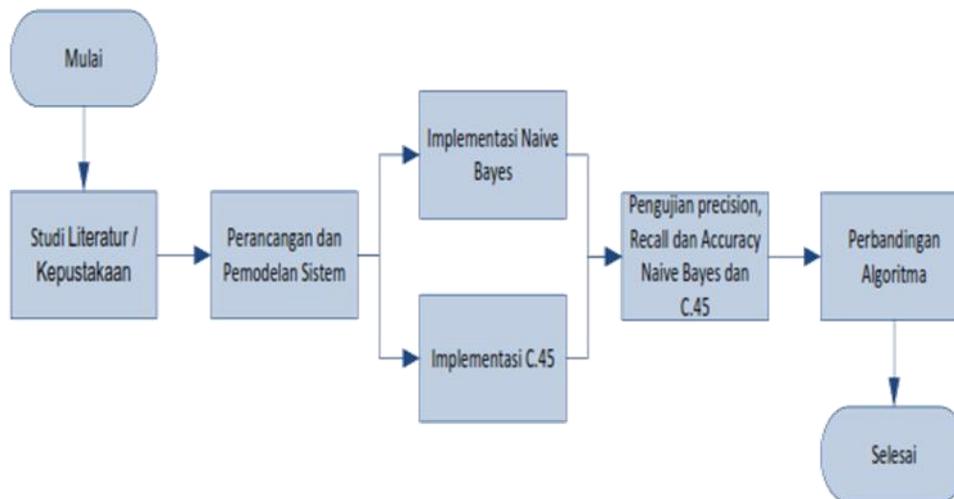
$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

$P(H)$ = Probabilitas hipotesis H (prior prob.)

$P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut $P(X)$ = Probabilitas dari X

METODE

Metode Penelitian yang diusulkan seperti terlihat pada bagan dibawah ini:



Gambar 1. Metode Penelitian

Bahan penelitian yang ditawarkan dalam penelitian ini adalah data preposisi, yaitu pengolahan data mentah dari penerimaan dana penelitian tahun 2015 sampai dengan tahun 2021 dengan jumlah dataset sebanyak 140. Data tersebut termasuk dalam pengumpulan data yang digunakan sebagai pemilihan data pelatihan. Setelah pengumpulan data, tahap selanjutnya adalah pembersihan data yang memiliki tahapan menentukan data yang sesuai dengan kriteria dan menghilangkan data yang tidak diperlukan. Kemudian menentukan kriteria yang akan digunakan dalam menentukan nilai probabilitas pada metode Naive Bayes dan Decision Tree.

Secara umum, penelitian ini terdiri dari dua tahap. Tahap pertama adalah persiapan atau preprocessing. Tahap pre-processing menyiapkan dataset asli berupa data excel di LPPM Universitas Tunas Pembangunan. Tahap kedua adalah proses, dimana pada tahap ini dilakukan pengolahan data untuk menemukan hasil yang diharapkan. Langkah-langkah yang dilakukan adalah sebagai berikut:

a. Implementasi Naive Bayes

Langkah yang dilakukan pada tahap ini adalah menerjemahkan desain yang telah terbentuk menjadi sistem yang menerapkan algoritma Naive Bayes yang menyatakan bahwa Naive Bayes Classifier merupakan model independen yang membahas klasifikasi sederhana berdasarkan teorema Bayes. Salah satu strategi yang dilakukan perusahaan adalah meningkatkan kualitas produk dan proses produksi yang dilakukan oleh perusahaan. Dalam dunia industri, kualitas dan produktivitas produk merupakan kunci keberhasilan proses produksi. Naive Bayes adalah algoritma yang dapat mengklasifikasikan variabel tertentu dengan menggunakan metode probabilitas dan statistik. Secara garis besar, algoritma Naive Bayes dapat dijelaskan sebagai berikut:

b. Implementasi Decision Tree

Decision Tree adalah mengubah data menjadi aturan keputusan. Manfaat utama dari penggunaan Decision Tree adalah kemampuannya untuk memecah proses pengambilan keputusan yang kompleks menjadi sederhana sehingga pengambilan keputusan akan memudahkan penyelesaian suatu masalah. Decision Tree digunakan untuk pengenalan pola dan disertakan dalam pengenalan pola statistik. Decision Tree menggunakan 2 perhitungan, yang pertama perhitungan Gain pada Persamaan 2 dan perhitungan Entropy pada Persamaan 3.

c. Pengujian Precision, Recall, dan Accuracy Naive Bayes dan Decision Tree



Menguji suatu algoritma membutuhkan standar dan alat uji. Membandingkan 2 algoritma harus memiliki standar yang sama agar dapat diketahui algoritma terbaik dari perbandingan tersebut. Pada tahap ini dilakukan pengujian dengan menghitung nilai presisi, recall, dan akurasi dari Naive Bayes dan Decision Tree. Langkah awal dalam tahapan ini adalah membagi data pada setiap kasus menjadi 2 yaitu data latih atau data latih dan data uji atau data uji. Data training digunakan sebagai data acuan dalam perhitungan masing-masing algoritma, sedangkan data testing digunakan untuk menilai prediksi dan penentuan yang dibuat oleh masing-masing algoritma sudah benar atau belum. Dalam membagi data menjadi data training dan data testing dilakukan beberapa perbandingan. Pada tahap ini dilakukan beberapa proses untuk mengolah data sebelum dilakukan klasifikasi menggunakan algoritma Naïve Bayes. Tahapan ini dilakukan karena preprocessing dapat meningkatkan akurasi Naive Bayes

HASIL DAN PEMBAHASAN

Setelah dilakukan beberapa tahapan penelitian mulai dari pre-processing atau persiapan sampai dengan proses yang meliputi beberapa tahapan antara lain pembersihan data, pengumpulan data, penentuan kriteria, penentuan probabilitas, dan pengujian, berikut adalah hasil dari penelitian tersebut:

Metode Naive Bayes

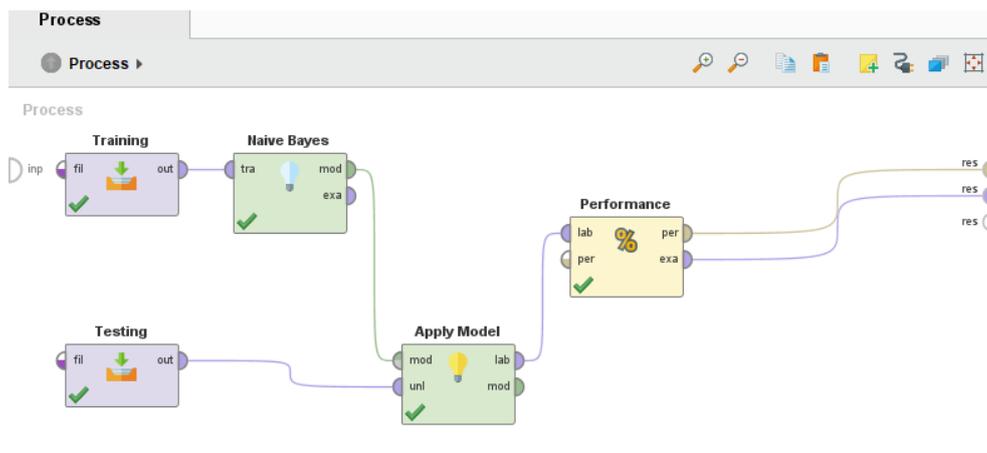
Tahapan penelitian yang akan dilakukan adalah data preposisi, yaitu mengolah data mentah dari penerimaan dana penelitian tahun 2015 sampai dengan tahun 2021 dengan total dataset sebanyak 140. Pada metode Naïve Bayes, data konstanta string atau kategorikal dibagi menjadi dua jenis, yaitu data numerik kontinu, sehingga akan terlihat perbedaannya pada saat menentukan nilai probabilitas dari setiap kriteria, baik kriteria dengan nilai data string maupun kriteria dengan data numerik.

Data tersebut kemudian siap untuk pelatihan dan pengujian dengan metode Naive Bayes Optimization dan pendekatan Diskritisasi Tanpa Pengawasan. Namun sebelum memasuki tahapan Naive Bayes Optimization menggunakan Rapidminer, terlebih dahulu metode Naive Bayes harus dilakukan dengan menggunakan rapidminer. Ada beberapa tahapan pengujian yang dilakukan yaitu mulai mengubah nilai Cross Validation dan Sampling Type yang dapat dilakukan dengan rapidminer. Untuk menguji dataset ini, dilakukan beberapa skenario penelitian seperti yang ditunjukkan pada tabel berikut:

Tabel 1. Uji Coba Validasi

Testing	Method	Number of Cross Validation	Sampling Type	Accuracy
1	Naïve Bayes	5	Automatic	89,53%
		5	Linear Sampling	87,39%
		5	Saffled Sampling	86,50%
		5	Stratified Sampling	82,60%
2	Naïve Bayes	10	Automatic	89,60%
		10	Linear Sampling	84,80%
		10	Saffled Sampling	87,20%
		10	Stratified Sampling	89,60%
3	Naïve Bayes	15	Automatic	89,81%
		15	Linear Sampling	85,60%
		15	Saffled Sampling	89,63%
		15	Stratified Sampling	89,81%

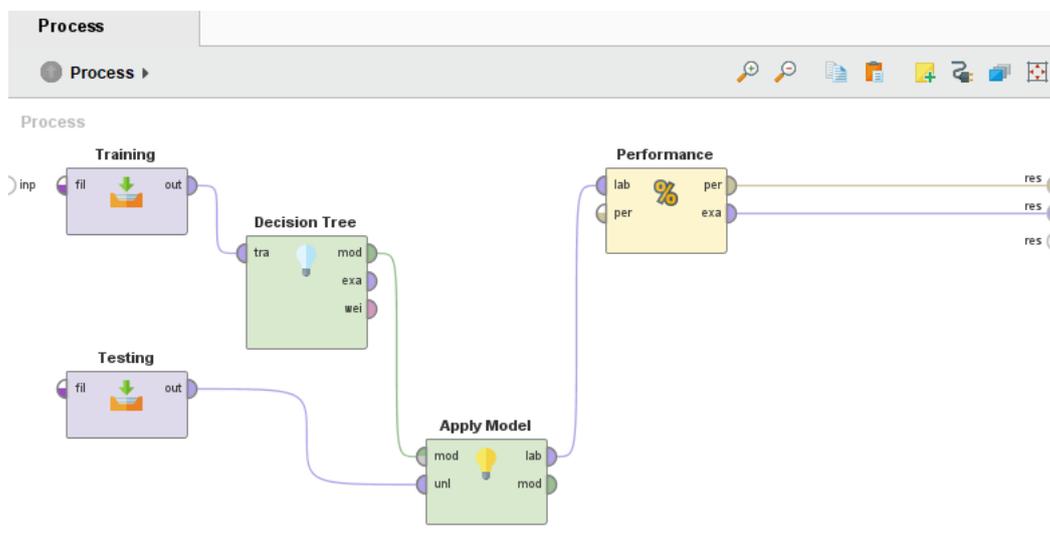
Dari beberapa pengujian diatas didapatkan nilai akurasi tertinggi yaitu 89,63%, nilai akurasi tersebut merupakan nilai yang tidak sesuai dengan yang diharapkan, karena menggunakan nilai Number Of Cross Validation 20 dan jenis pengambilan sampel menggunakan Automatic .



Gambar 2. Desain Rapid Miner Naïve Bayes

Selanjutnya dataset yang digunakan dibagi menjadi dua bagian dengan perbandingan 70% digunakan untuk data training dengan 100 record data dan 30% untuk data testing dengan 25 record data dari total jumlah record dari 125 dataset. Dan hasil akurasi yang didapatkan dari desain eksperimen dengan menggunakan metode Naive Bayes adalah nilai akurasi sebesar 96%.

Rancangan eksperimen di atas menggambarkan bahwa dataset yang digunakan terbagi menjadi dua bagian dengan 70% digunakan untuk data training dengan 100 record data dan 30% data testing dengan 25 record data dari total jumlah record dari 125 dataset. Dan hasil akurasi yang didapatkan dari desain eksperimen di atas dengan menggunakan metode Naive Bayes adalah nilai akurasi sebesar 96%. Dalam pengujian dengan menggunakan metode Decision Tree, ada beberapa kriteria yang digunakan antara lain gain_ratio, information_gain, gini_index, dan akurasi. Kriteria ini adalah salah satu dari beberapa operator yang akan menghasilkan perkiraan seberapa akurat sebuah model akan terlihat. Rancangan pengujian dengan menggunakan metode Decision Tree adalah sebagai berikut:



Gambar 3. Desain Rapid Miner Decision Tree

Rancangan eksperimen di atas menggambarkan bahwa dataset yang digunakan dibagi menjadi dua bagian dengan perbandingan 70% digunakan untuk data training dengan 100 record data dan 30% untuk data testing dengan 25 record data dari total jumlah record dari 125 dataset. Pengujian data training dan testing di atas dengan menggunakan metode Decision Tree diperoleh akurasi sebesar 64%. Nilai akurasi ini diperoleh dengan berbagai pengujian, diantaranya kriteria bernama gini_index dengan nilai kedalaman maksimal 20, Apply Pruning Confidence 0.1 dan 0.2, serta Apply Prepruning 0.01 dan 0.02. Jika pengujian



dilakukan dengan nilai diatas nilai kedalaman maksimal 20, misal 50 dan nilai Apply Pruning Confidence maka nilai Apply Prepruning sudah fix maka hasil yang didapat untuk nilai juga sama yaitu 64% .

KESIMPULAN DAN SARAN

Dari pembahasan di atas, dapat ditarik beberapa kesimpulan, antara lain:

1. Pengklasifikasian penentuan program studi untuk pendanaan proposal penelitian dapat dilakukan dengan beberapa metode, antara lain metode Naïve Bayes dan Decision Tree.
2. Hasil penelitian yang dilakukan dari tahap awal hingga pengujian dengan menerapkan metode klasifikasi Naïve Bayes diperoleh akurasi sebesar 96%, dan metode Decision Tree diperoleh akurasi sebesar 64%.
3. Hasil akhir metode klasifikasi terbaik dalam penelitian ini untuk menentukan pendanaan proposal penelitian adalah metode Naïve Bayes, karena nilai akurasi yang diperoleh sebesar 96%.

DAFTAR PUSTAKA

- Ahmed, M. S., Shahjaman, M., Kabir, E., & Kamruzzaman, M. (2018). Prediction of protein acetylation sites using kernel naive Bayes classifier based on protein sequences profiling. *Bioinformation*, 14(5), 213.
- Aprilla, D., Baskoro, D. A., Ambarwati, L., & Wicaksana, I. W. S. (2013). Belajar data mining dengan rapidminer. *Jakarta: Gramedia Pustaka Utama*.
- Budiyantara, A., Irwansyah, I., Prengki, E., Pratama, P. A., Wiliani, N., & others. (2020). Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus Tepat Waktu. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 5(2), 265–270.
- Buslim, N., & Iswara, R. P. (2019). Pengembangan Algoritma Unsupervised learning technique pada big data analysis di media sosial sebagai media promosi online bagi masyarakat. *Jurnal Teknik Informatika*, 12(1), 79–96.
- Hardiani, T. (2021). Comparison of Naive Bayes Method, K-NN (K-Nearest Neighbor) and Decision Tree for Predicting the Graduation of 'Aisyiyah University Students of Yogyakarta. *Int. J. Heal. Sci. Technol*, 2(1).
- Hozairi, H., Anwari, A., & Alim, S. (2021). Implementasi Orange Data Mining Untuk Klasifikasi Kelulusan Mahasiswa Dengan Model K-Nearest Neighbor, Decision Tree Serta Naive Bayes. *Network Engineering Research Operation*, 6(2), 133–144.
- Jananto, A. (2013). Algoritma Naive Bayes untuk Mencari Perkiraan Waktu Studi Mahasiswa. *Dinamik*, 18(1).
- Maswadi, K., Ghani, N. A., Hamid, S., & Rasheed, M. B. (2021). Human activity classification using Decision Tree and Naive Bayes classifiers. *Multimedia Tools and Applications*, 80(14), 21709–21726.
- Mustafa, M. S., Ramadhan, M. R., & Thenata, A. P. (2018). Implementasi Data Mining untuk Evaluasi Kinerja Akademik Mahasiswa Menggunakan Algoritma Naive Bayes Classifier. *Creative Information Technology Journal*, 4(2), 151–162.
- Nurajijah, N., & Riana, D. (2019). Algoritma naive Bayes, decision tree, dan SVM untuk klasifikasi persetujuan pembiayaan nasabah koperasi syariah. *Jurnal Teknologi Dan Sistem Komputer*, 7(2), 77–82.
- Oktasari, L., Chrisnanto, Y. H., & Yuniarti, R. (2016). Text Mining Dalam Analisis Sentimen Asuransi Menggunakan Metode Naive Bayes Classifier. *Prosiding SNST Fakultas Teknik*, 1(1).
- Panda, M., Abraham, A., & Patra, M. R. (2010). Discriminative multinomial naive bayes for network intrusion detection. *2010 Sixth International Conference on Information Assurance and Security*, 5–10.



- Rahman, M. A., & Akter, Y. A. (2019). Topic classification from text using decision tree, K-NN and multinomial naive bayes. *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, 1–4.
- Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, 7(2), 156–165.
- Romadloni, N. T., Santoso, I., & Budilaksono, S. (2019). Perbandingan Metode Naive Bayes, KNN dan Decision Tree Terhadap Analisis Sentimen Transportasi KRL Commuter Line. *Ikrath-Informatika*, 3(2), 1–9.
- Saritas, M. M., & Yasar, A. (2019). Performance analysis of ANN and Naive Bayes classification algorithm for data classification. *International Journal of Intelligent Systems and Applications in Engineering*, 7(2), 88–91.
- Sindu, V., & Prakash, M. (2021). Comparative Analysis Among Decision Tree vs. Naive Bayes for Prediction of Weather Prognostication. *Smart Intell. Comput. Commun. Technol*, 38, 94–100.
- Singh, N., & Singh, P. (2019). A novel Bagged Naive Bayes-Decision Tree approach for multi-class classification problems. *Journal of Intelligent & Fuzzy Systems*, 36(3), 2261–2271.
- Tripathi, A., Yadav, S., & Rajan, R. (2019). Naive Bayes classification model for the student performance prediction. *2019 2nd International Conference on Intelligent Computing, Instrumentation and Control Technologies (ICICT)*, 1, 1548–1553.
- Westphal, P., Böhmann, L., Bin, S., Jabeen, H., & Lehmann, J. (2019). SML-Bench—A benchmarking framework for structured machine learning. *Semantic Web*, 10(2), 231–245. <https://dl.acm.org/doi/abs/10.3233/SW-180308>
- Yadav, K., & Thareja, R. (2019). Comparing the performance of naive bayes and decision tree classification using R. *International Journal of Intelligent Systems and Applications*, 11(12), 11.