BAB V
PENUTUP

5.1 Kesimpulan

Berdasarkan rangkaian eksperimen, pengujian, dan analisis data yang dilakukan menggunakan perangkat lunak RapidMiner terhadap data calon Kelompok Usaha Bersama (KUBE) Anggaran Tahun 2026 di Dinas Sosial Pengendalian Penduduk dan Keluarga Berencana Kabupaten Purworejo, telah diperoleh gambaran komprehensif mengenai performa algoritma K-Nearest Neighbors (K-NN) dan Naïve Bayes. Penelitian ini mengevaluasi kelayakan calon Kelompok Usaha Bersama menggunakan tiga skenario pembagian data latih dan data uji, yaitu 20:80, 50:50, dan 80:20. Hasil evaluasi menunjukkan bahwa proporsi pembagian data memberikan pengaruh signifikan terhadap tingkat *accuracy*, *precision*, *recall*, serta *f-score* dari masing-masing model klasifikasi yang diuji.

Pengujian menggunakan algoritma K-Nearest Neighbors (K-NN) dilakukan dengan tiga skenario pembagian data, yaitu 20:80, 50:50, dan 80:20. Pada skenario pembagian data 20:80, dengan 20% data sebagai data latih dan 80% sebagai data uji, K-NN memperoleh *accuracy* sebesar 97,14% dan *f-score* sebesar 97,05%. Selanjutnya, pada skenario 50:50, *accuracy* K-NN mengalami penurunan menjadi 94,32%, sedangkan *f-score* juga menurun menjadi 94,19%. Penurunan tersebut menunjukkan bahwa perubahan proporsi data pada skenario 50:50 belum memberikan peningkatan kinerja bagi K-NN. Sementara itu, pada skenario pembagian data 80:20, K-NN

menunjukkan performa terbaik dengan memperoleh accuracy dan *f-score* sebesar 100,00%. Hasil tersebut menunjukkan bahwa penggunaan 80% data sebagai data latih memberikan hasil klasifikasi yang optimal pada K-NN.

Pada pengujian menggunakan algoritma Naïve Bayes, hasil yang diperoleh menunjukkan performa yang relatif stabil pada ketiga skenario pembagian data. Pada skenario 20:80, Naïve Bayes memperoleh *accuracy* sebesar 96,43% dan *f-score* sebesar 96,51%. Pada skenario 50:50, *accuracy* meningkat menjadi 96,59%, sedangkan *f-score* sedikit menurun menjadi 96,48%. Perubahan tersebut menunjukkan bahwa meskipun terjadi peningkatan *accuracy*, nilai *f-score* Naïve Bayes relatif tetap. Selanjutnya, pada skenario 80:20, Naïve Bayes mencapai performa terbaik dengan *accuracy* dan *f-score* sebesar 100,00%. Dengan demikian, seperti halnya K-NN, penggunaan proporsi data latih sebesar 80% memberikan hasil klasifikasi paling optimal bagi Naïve Bayes.

Secara keseluruhan, hasil pengujian menunjukkan adanya perbedaan kinerja antara K-NN dan Naïve Bayes pada skenario 20:80 dan 50:50, sedangkan pada skenario 80:20 kedua algoritma menghasilkan kinerja yang sama. Pada skenario 20:80, K-NN memiliki *accuracy* 97,14% dan *f-score* 97,05%, lebih tinggi dibandingkan Naïve Bayes yang memperoleh accuracy 96,43% dan *f-score* 96,51%. Sebaliknya, pada skenario 50:50, Naïve Bayes menunjukkan hasil yang lebih baik dengan *accuracy* 96,59% dan *f-score* 96,48%, sedangkan K-NN memperoleh *accuracy* 94,32% dan *f-score* 94,19%. Pada skenario 80:20, kedua metode memperoleh *accuracy* dan *f-*

score sebesar 100,00%, sehingga tidak terdapat perbedaan kinerja berdasarkan kedua metrik tersebut.

Berdasarkan hasil tersebut, tidak dapat disimpulkan bahwa salah satu algoritma secara keseluruhan selalu lebih baik daripada algoritma lainnya, karena hasilnya bergantung pada skenario pembagian data yang digunakan. K-NN memberikan hasil lebih tinggi pada skenario 20:80, sedangkan Naïve Bayes memberikan hasil lebih tinggi pada skenario 50:50. Namun, kedua algoritma menunjukkan performa terbaik dan hasil yang sama pada skenario 80:20, sehingga skenario tersebut dapat dipandang sebagai pembagian data yang paling optimal dalam pengujian ini. Selain itu, Naïve Bayes menunjukkan perubahan *accuracy* yang lebih kecil antara skenario 20:80 dan 50:50 dibandingkan K-NN, sehingga pada kondisi dengan proporsi data latih yang belum mencapai 80%, Naïve Bayes cenderung memberikan hasil yang lebih stabil.

5.2 Saran

Penelitian ini masih memiliki beberapa keterbatasan, terutama pada cakupan variasi data yang digunakan dalam proses pengujian model klasifikasi. Oleh karena itu, penelitian selanjutnya disarankan menggunakan dataset yang memiliki atribut atau kriteria yang lebih beragam sehingga mampu merepresentasikan karakteristik data yang lebih kompleks. Pada penelitian ini, algoritma K-Nearest Neighbors (KNN) dan Naive Bayes menghasilkan nilai *accuracy* yang sama, yaitu sebesar 100% pada skenario pembagian data 80:20. Kondisi tersebut menunjukkan bahwa kedua algoritma

memiliki performa yang relatif setara terhadap dataset yang digunakan. Dengan menggunakan data yang lebih beragam, penelitian selanjutnya diharapkan dapat mengevaluasi sensitivitas masing-masing algoritma, mengetahui batas konsistensi performanya, serta memperoleh gambaran yang lebih komprehensif mengenai keunggulan dan keterbatasan setiap metode ketika diterapkan pada data dengan karakteristik yang lebih beragam.

Apabila penelitian selanjutnya masih mengangkat topik mengenai klasifikasi calon Kelompok Usaha Bersama, disarankan untuk memperluas dan memperdalam kriteria seleksi yang digunakan. Selain memanfaatkan aspek administratif, penilaian dapat dikembangkan dengan menambahkan indikator yang lebih representatif terhadap kondisi asli calon anggota kelompok, seperti kondisi sosial ekonomi dan finansial secara lebih mendalam, riwayat atau pengalaman usaha, jenis dan sektor usaha produktif yang dijalankan, kapasitas sumber daya manusia dalam kelompok, serta tingkat komitmen dan kesiapan manajerial dalam mengelola bantuan modal usaha. Penambahan atribut tersebut diharapkan mampu meningkatkan kualitas proses klasifikasi sehingga hasil yang diperoleh tidak hanya mencerminkan kelayakan administratif, tetapi juga potensi keberlanjutan dan keberhasilan usaha kelompok di masa mendatang.

Aspek prapemrosesan data juga perlu menjadi perhatian utama pada penelitian berikutnya. Pada penelitian ini, jumlah data yang digunakan mengalami pengurangan karena terdapat data yang tidak lengkap, khususnya pada atribut desil yang tidak tersedia atau tidak ditemukan pada Data Terpadu

Kesejahteraan Sosial (DTKS) maupun Data Terpadu Sosial Ekonomi Nasional (DTSEN). Kondisi tersebut menyebabkan sebagian data tidak dapat digunakan dalam proses klasifikasi. Oleh karena itu, peneliti selanjutnya disarankan melakukan validasi dan sinkronisasi data secara lebih intensif dengan pihak Dinas Sosial sebelum proses analisis dilakukan. Selain itu, penerapan teknik penanganan missing value, seperti metode imputation, dapat dipertimbangkan agar jumlah data yang dapat dimanfaatkan tetap optimal tanpa mengurangi kualitas analisis.

Dari sisi metodologi, penelitian selanjutnya juga dapat mengembangkan skenario pengujian yang lebih luas. Pada algoritma K-Nearest Neighbors, misalnya, pengujian dapat dilakukan dengan menggunakan variasi nilai tetangga terdekat yang lebih beragam serta membandingkan beberapa metrik jarak, seperti *Euclidean Distance*, *Manhattan Distance*, maupun *Minkowski Distance*, untuk mengetahui pengaruhnya terhadap performa klasifikasi.

Terakhir, hasil penelitian ini diharapkan tidak hanya berhenti pada tahap evaluasi performa model menggunakan *Confusion Matrix* melalui perangkat lunak RapidMiner. Penelitian selanjutnya dapat mengembangkan model klasifikasi terbaik ke dalam sebuah sistem pendukung keputusan atau aplikasi berbasis web yang dapat digunakan secara langsung oleh Dinas Sosial. Implementasi tersebut diharapkan mampu membantu proses seleksi calon Kelompok Usaha Bersama secara lebih cepat, objektif, transparan, konsisten, dan tepat sasaran, sehingga dapat meningkatkan efektivitas

pengambilan keputusan serta mendukung optimalisasi penyaluran bantuan sosial kepada kelompok yang benar-benar memenuhi kriteria.