BAB V

PENUTUP

5.1. Kesimpulan

1. Hasil dari proses clustering menggunakan algoritma K-Means menunjukkan bahwa data dapat dikelompokkan ke dalam tiga cluster utama, yang masing-masing menggambarkan pola kemiripan atribut tertentu di dalam data. Ketiga cluster ini mengelompokkan data berdasarkan kemiripan fitur tanpa memperhatikan label target, sehingga berfungsi sebagai cara eksplorasi awal untuk melihat bagaimana data tersebar secara alami. Meskipun hasil cluster tidak secara langsung menunjukkan kategori seperti puas atau tidak puas (dalam konteks survei atau kepuasan misalnya), namun dapat digunakan sebagai referensi untuk memetakan pola atau karakteristik umum yang muncul dalam data. Clustering ini juga dapat membantu proses validasi internal, apakah hasil klasifikasi nantinya memiliki pola yang sejalan dengan distribusi alami data.
2. Setelah dilakukan klasifikasi menggunakan algoritma Naive Bayes terhadap 1000 data, diperoleh distribusi prediksi sebanyak 292 data pada kelas C1, 419 data pada C2, dan 289 data pada C3. Hasil klasifikasi ini bersifat supervised, artinya data telah memiliki label kategori target, dan proses klasifikasi dilakukan untuk memprediksi dengan akurat kelas yang sesuai berdasarkan input fitur yang diberikan. Jika dibandingkan secara tidak langsung dengan hasil clustering, dapat diamati bahwa ada kemiripan dalam jumlah distribusi antar kelompok, yang mengindikasikan bahwa pemisahan

data yang dilakukan baik oleh K-Means maupun Naive Bayes memiliki keteraturan dan konsistensi dalam mengenali pola data. Hal ini memperkuat bahwa struktur data cukup stabil dan model machine learning mampu bekerja secara efektif dalam mengenali karakteristiknya.

3. Evaluasi performa model Naive Bayes melalui confusion matrix dan metrik evaluasi menunjukkan hasil yang sangat baik. Dari confusion matrix, terlihat bahwa sebagian besar data diklasifikasikan dengan benar: 275 dari 292 data C1, 404 dari 419 data C2, dan 238 dari 289 data C3, dengan kesalahan klasifikasi yang cukup kecil. Sementara itu, metrik evaluasi juga mengonfirmasi performa tinggi dengan nilai AUC sebesar 0.990, akurasi 0.917, precision 0.920, recall 0.917, dan F1-score 0.916. Hal ini menunjukkan bahwa model memiliki sensitivitas dan presisi yang seimbang, serta mampu mengklasifikasikan data secara andal. Keseluruhan hasil ini menegaskan bahwa integrasi antara analisis awal dengan K-Means dan klasifikasi lanjutan menggunakan Naive Bayes dapat menjadi kombinasi yang kuat dalam eksplorasi serta prediksi data berbasis kategori.

5.2. Saran

Berdasarkan hasil pengelompokan data menggunakan algoritma K-Means, disarankan agar peneliti selanjutnya melakukan analisis lebih dalam terhadap karakteristik setiap cluster. Misalnya, dengan mengidentifikasi fitur-fitur dominan yang membentuk masing-masing kelompok, dapat diketahui kecenderungan tertentu dalam data yang mungkin relevan untuk pengambilan keputusan atau strategi bisnis. Selain itu, penggunaan jumlah cluster yang berbeda, seperti K=2

atau K=4, juga bisa diuji untuk melihat apakah ada pemisahan data yang lebih optimal. Dengan pendekatan ini, hasil clustering tidak hanya menjadi alat eksploratif, tetapi juga mendukung proses klasifikasi secara lebih tajam melalui pemahaman pola distribusi data yang lebih menyeluruh.

Dalam hal klasifikasi, algoritma Naive Bayes telah menunjukkan performa yang sangat baik dengan akurasi dan nilai AUC yang tinggi. Namun, untuk penelitian lanjutan, disarankan untuk membandingkan kinerja Naive Bayes dengan algoritma klasifikasi lainnya seperti Decision Tree, Support Vector Machine (SVM), atau Random Forest. Perbandingan ini penting agar diperoleh pemodelan yang paling sesuai dengan karakteristik data. Selain itu, penggunaan teknik validasi silang yang lebih kompleks, seperti k-fold cross validation dengan jumlah lipatan lebih besar, dapat menghasilkan evaluasi model yang lebih stabil dan akurat, serta menghindari kemungkinan overfitting pada data.

Selanjutnya, agar hasil klasifikasi semakin bermanfaat secara praktis, saran diberikan untuk mengintegrasikan sistem klasifikasi ini ke dalam aplikasi atau sistem informasi yang dapat digunakan oleh instansi atau pihak terkait. Misalnya, sistem ini dapat membantu dalam pengambilan keputusan berbasis prediksi terhadap perilaku atau kategori data tertentu. Di samping itu, penting pula untuk memperhatikan kualitas dan keseimbangan data latih di masa mendatang, karena data yang tidak seimbang dapat memengaruhi performa model. Dengan pengumpulan data yang lebih luas dan representatif, sistem klasifikasi yang dikembangkan akan menjadi lebih kuat, adaptif, dan relevan dalam konteks implementasi nyata.