

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil dari penelitian dan pembahasan, kesimpulan yang dapat diberikan adalah:

1. Terdapat hasil analisa kebutuhan terhadap klasifikasi teks untuk mengklasifikasikan konten website yang mengandung unsur pornografi atau tidak. Kebutuhan tersebut adalah adanya situs dengan konten porno di dalamnya yang dapat diakses oleh anak usia belajar yang tidak diketahui oleh orang tua, guru, dan masyarakat umum lainnya dalam KBM berbasis daring.
2. Model klasifikasi dengan menggunakan algoritma *Gaussian Naïve Bayes*, *Bernoulli Naïve Bayes*, dan *Multinomial Naïve Bayes*. Untuk menentukan model algoritma yang paling tepat digunakan, dilakukan pengujian pada ketiga algoritma tersebut dan dipilih model dengan nilai akurasi tertinggi. Selain akurasi, model juga harus mencakup matrik evaluasi kinerja lainnya seperti presisi, *recall*, dan *F1-score* untuk memastikan kinerjanya yang baik. Dalam hal ini model *Multinomial Naïve Bayes* merupakan model klasifikasi yang memiliki performa atau kinerja model yang lebih baik.
3. Terdapat hasil rancangan tahapan atau proses untuk mengklasifikasikan konten *website* yang mengandung unsur pornografi dan non-pornografi menggunakan algoritma *Naïve Bayes* dengan menggunakan model *Multinomial Naïve Bayes* yaitu dengan melalui tahap proses seperti *model requirements*, *data collection*, *data cleaning*, *data labeling*, *feature engineering*, *model training* dan *model evaluation*.

4. Model klasifikasi *Multinomial Naïve Bayes* yang digunakan dalam penelitian ini memiliki performa yang sangat baik dengan nilai akurasi awal sebesar 93.38%. Namun, setelah dilakukan *tuning hyperparameter*, *report classification*, dan pencegahan *overfitting* dan *underfitting*, nilai akurasi model mengalami peningkatan menjadi 93.51%. Selanjutnya, penelitian juga menggunakan beberapa teknik untuk melihat performa model dengan menggunakan *5 fold-validation* dan didapatkan hasil akhir akurasi model klasifikasi sebesar 94.61% serta nilai kurva ROC yang sangat tinggi yaitu 98.73% (*Excellent Classification*). Hal ini menunjukkan bahwa model klasifikasi *Multinomial Naïve Bayes* dapat digunakan untuk memprediksi apakah suatu konten pada *website* mengandung unsur pornografi atau tidak dengan akurasi yang tinggi.
5. Model klasifikasi *Multinomial Naïve Bayes* yang telah dibuat dalam penelitian ini dapat mengklasifikasikan teks dalam kategori pornografi dan non-pornografi dengan tepat dan akurat. Dalam penelitian ini, model tersebut telah diuji dengan membandingkan hasil klasifikasi berdasarkan pendapat dari ahli bahasa Indonesia dan ditemukan bahwa model *Multinomial Naïve Bayes* dapat mengklasifikasikan teks tersebut dengan sesuai dan tepat. Hal ini menunjukkan bahwa model klasifikasi yang dibangun dapat digunakan untuk mengklasifikasikan konten pada *website* dan membantu mengidentifikasi konten pornografi dengan akurasi yang tinggi.

5.2 Saran

Saran yang dapat di berikan dari hasil penelitian yang sudah dilakukan adalah Untuk pengembangan penelitian selanjutnya, berikut adalah penjelasan untuk setiap kelemahan sistem yang telah diidentifikasi:

1. *Dataset* yang tidak seimbang dapat menyebabkan bias dalam model dan hasil prediksinya. Oleh karena itu, untuk mengatasi kelemahan ini, penelitian selanjutnya dapat menggunakan teknik sampling atau weighting pada *dataset* untuk memastikan keseimbangan antara kelas positif dan negatif. Selain itu, perlu juga dilakukan evaluasi terhadap kinerja model menggunakan matrik lain seperti *precision*, *f1-score*, dan *ROC-AUC*, yang dapat memberikan gambaran yang lebih lengkap tentang performa model pada *dataset* yang tidak seimbang.
2. Penggunaan parameter *alpha* yang tidak seimbang pada *validation curve* dapat mengakibatkan *overfitting* atau *underfitting* pada model. Oleh karena itu, penelitian selanjutnya dapat melakukan eksplorasi pada rentang nilai *alpha* yang lebih luas dan melakukan pengecekan terhadap performa model pada setiap nilai *alpha* tersebut. Selain itu, juga perlu dilakukan pengujian pada metode lain dalam model machine learning yang dapat memberikan performa yang lebih baik pada *dataset* yang digunakan.
3. Nilai *recall* kelas positif sebesar 85.95% menunjukkan bahwa model cenderung memprediksi *values negative* pada kelas positif. Untuk meningkatkan nilai *recall* pada kelas positif, penelitian selanjutnya dapat melakukan teknik *feature engineering* lainnya yang berbeda pada *dataset* untuk meningkatkan pemisahan antara kelas positif dan negatif. Selain itu, penggunaan model yang lebih kompleks seperti model *ensemble* atau *deep learning* dapat juga dipertimbangkan untuk meningkatkan performa model dalam memprediksi kelas positif serta berdasarkan analisa yang telah

dilakukan, peneliti dapat menarik permasalahan utama dalam prediksi kelas positif ini adalah dimana ketidakseimbangan antara kedua label kelas yang memberikan pengaruh besar sehingga prediksi kelas positif memiliki performa yang kurang maksimal. Ketidakseimbangan dataset yang dimaksud adalah jumlah label kelas pada proses labeling data.

4. Untuk pengembangan model klasifikasi *Multinomial Naïve Bayes* ada baiknya jika model yang dibuat dapat diimplementasikan dalam bentuk aplikasi atau *website* sehingga memudahkan *user* atau pengguna dalam mengklasifikasikan dokumen *website*.