BAB V

SIMPULAN DAN SARAN

5.1. Kesimpulan

Maraknya komentar spam promosi judi online di YouTube mendorong pengumpulan data menggunakan *YouTube Data API v3* dari 20 channel, masing-masing 50 video terbaru, menghasilkan 1.071.028 baris komentar mentah. Setelah penghapusan duplikat tersisa 897.894 komentar unik, dengan distribusi kelas yang sangat timpang (880.711 komentar "biasa" vs 17.183 komentar promosi judi). Untuk mengurangi bias, dilakukan *random undersampling* sehingga masing-masing kelas berjumlah sekitar 14.000 sampel, lalu teks dikonversi ke vektor *TF-IDF* berdimensi 5.000.

Tahap *benchmarking* terhadap sebelas algoritma menunjukkan bahwa K-Nearest Neighbors (KNN) default mencapai *accuracy* 98,48 %, *precision* 78,03 %, *recall* 90,52 %, dan *F1-score* makro 83,07 %, mengungguli Logistic Regression (LR), CatBoost, LightGBM, XGBoost, dan Linear SVM. Pada hyperparameter tuning, KNN diubah menjadi *algorithm='ball_tree', leaf_size=42, n_neighbors=2, p=2, weights='distance'*, menghasilkan *recall* naik ke 94 % tetapi *precision* turun ke 71 %, *accuracy* ke 97 %, dan *F1-score* ke 78 % (runtime 139,5 s). Sementara itu, LR dengan *C=5.95, class_weight='balanced', penalty='l2',* dan *solver='sag'* menaikkan *recall* ke 97 % dengan *precision* 74 %, *accuracy* 98 %, *F1-score* 81 % (runtime 0,05 s).

Analisis trade-off menunjukkan bahwa meski peningkatan *recall* KNN lebih besar (+3,48 % vs +0,47 %), penurunan *precision* (−7,03 % vs −1,40 %) dan *F1-score* (−5,07 % vs −1,61 %) jauh lebih dalam daripada LR. Dengan prioritas meminimalkan *false negatives* sambil mempertahankan *precision* dan *accuracy*, LR ter-tuning terpilih sebagai model akhir. Model ini siap diintegrasikan sebagai moderasi komentar otomatis dan *early-warning system* bagi moderator manusia.

5.2. Saran

Kendati kontribusi yang diberikan, masih terdapat berbagai keterbatasan yang membuka ruang pengembangan lebih lanjut. Metodologi dan temuan yang dihasilkan diharapkan dapat menjadi dasar bagi studi-studi selanjutnya—terutama dalam ranah akademik yang memanfaatkan dokumen tidak terstruktur. Untuk itu, beberapa rekomendasi pengembangan di masa depan meliputi:

1. **Pendekatan Berbasis *Deep Learning*:** Eksperimen dengan *fine-tuning* model transformer Bahasa Indonesia (misalnya IndoBERT atau XLM-R) untuk menangkap konteks lebih dalam daripada TF-IDF. Arsitektur hybrid CNN-LSTM dengan mekanisme attention dapat memperkuat ekstraksi pola lokal dan urutan kata. *Multi-task learning*—menggabungkan klasifikasi topik dan sentimen—juga bisa meningkatkan generalisasi.
2. **Pengembangkan *Ensemble Learning:*** Penggunaan *stacking* atau *blending* untuk menggabungkan kekuatan model seperti XGBoost, CatBoost, dan CNN, lalu padukan dengan *meta-learner* (misalnya Logistic Regression). *Dynamic ensemble selection* memungkinkan pemilihan *sub-ensemble* terbaik berdasarkan karakteristik setiap sampel. Pelatihan *adversarial* dan *regularisasi* pada level *ensemble* dapat mengurangi *overfitting*.
3. Penambahan Attribute: Penambahan fitur linguistik seperti *part-of-speech tagging*, *dependency parsing*, dan *named-entity recognition* untuk menangkap struktur sintaksis. Sertakan topic modeling (LDA) atau *sentence embedding* (Sentence-BERT) agar tema diskusi lebih terdeteksi. Metadata—misalnya waktu posting dan jumlah interaksi—juga dapat membantu membedakan komentar promosi judi dari komentar biasa.