

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa terdapat perbedaan pola penggunaan bahasa antara pengguna laki-laki dan perempuan pada teks cuitan media sosial X. Analisis kata dominan dan kata khas menunjukkan bahwa pengguna laki-laki cenderung menggunakan kata yang berkaitan dengan aktivitas, kompetisi, topik teknis, serta konteks informatif. Sementara itu, pengguna perempuan lebih banyak menggunakan kata yang bersifat ekspresif, interpersonal, serta berkaitan dengan perasaan dan pengalaman personal. Temuan ini menunjukkan bahwa karakteristik linguistik dapat dimanfaatkan sebagai indikator yang relevan dalam membedakan gender pengguna berbasis teks, sekaligus menegaskan bahwa analisis bahasa memiliki peran penting dalam proses klasifikasi gender.

Selain itu, hasil pengujian menunjukkan adanya perbedaan performa antara algoritma Logistic Regression, Random Forest, dan Support Vector Machine (SVM) dalam melakukan klasifikasi gender berbasis teks cuitan. Dari ketiga algoritma tersebut, SVM memberikan hasil terbaik dengan rata-rata akurasi sebesar 89,76% serta simpangan baku sebesar 0,0276. Hasil ini menunjukkan bahwa SVM linear mampu memanfaatkan representasi fitur TF-IDF dengan baik, khususnya pada data teks yang berdimensi tinggi dan bersifat sparse. Dengan demikian, dapat disimpulkan bahwa algoritma SVM linear merupakan metode yang paling optimal untuk digunakan dalam klasifikasi gender berbasis teks pada penelitian ini.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya:

1. Penelitian selanjutnya dapat menggunakan dataset dengan jumlah yang lebih besar atau melibatkan variasi bahasa yang lebih beragam guna meningkatkan kemampuan generalisasi model.
2. Penelitian berikutnya juga dapat mengeksplorasi metode representasi teks berbasis word embedding modern seperti Word2Vec, FastText, atau model

berbasis transformer seperti BERT untuk membandingkan performanya dengan metode TF-IDF yang digunakan pada penelitian ini.

3. Selain itu, penggunaan algoritma deep learning atau pendekatan hybrid dapat dipertimbangkan untuk menangkap pola bahasa yang lebih kompleks pada data teks media sosial.
4. Penelitian selanjutnya juga perlu mempertimbangkan aspek etika dan potensi bias dalam klasifikasi gender, mengingat gender merupakan karakteristik yang sensitif dan memiliki kompleksitas sosial.

