

BAB 5 SIMPULAN SARAN

5.1 Simpulan

Pengembangan model dilakukan dengan *Knowledge Discovery in Databases (KDD)* yang melibatkan tahapan mulai dari pengumpulan data, pembersihan, transformasi, hingga evaluasi model sebagai metode penelitian. Berdasarkan analisis dan perbandingan yang dilakukan antara model *Logistic Regression* dan *Ridge Classifier* untuk klasifikasi berita hoaks dalam bahasa Indonesia, dapat disimpulkan bahwa kedua model tersebut menunjukkan performa yang sangat baik dan sebanding. Dengan capaian akurasi 98% untuk *Logistic Regression* dan 97% untuk *Ridge Classifier*. Meskipun terdapat perbedaan tipis pada beberapa metrik evaluasi, seperti akurasi, presisi, *recall*, dan *F1-score*, kedua model mencapai hasil yang memuaskan dengan nilai metrik yang tinggi. Evaluasi menggunakan teknik *cross-validation* dengan *5 fold* dan *StratifiedKFold* memberikan visualisasi yang menampilkan model *Logistic Regression* lebih stabil dibandingkan dengan *Ridge Classifier*.

Meskipun demikian, penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Model saat ini mengandalkan fitur tekstual dan pola linguistik untuk mengklasifikasikan berita, tanpa mempertimbangkan akurasi faktual dari konten berita. Hal ini dapat menyebabkan model salah mengklasifikasikan berita hoaks yang memiliki karakteristik bahasa yang serupa dengan berita faktual. Selain itu, *dataset* yang digunakan memiliki ketidakseimbangan, dengan artikel hoaks yang cenderung lebih pendek dibandingkan artikel faktual. Ketidakseimbangan ini dapat mempengaruhi kemampuan model dalam menangani berita dengan jumlah kata yang bervariasi.

5.2 Saran

Untuk penelitian selanjutnya, beberapa saran yang dapat dipertimbangkan antara lain:

1. Meninvestigasi teknik-teknik *preprocessing* yang lebih canggih, seperti normalisasi teks, penanganan *slang* atau singkatan, atau pemodelan topik, untuk meningkatkan kualitas representasi fitur.

2. Mengintegrasikan mekanisme pemeriksaan fakta ke dalam model klasifikasi. Dengan mempertimbangkan akurasi faktual dari konten berita, model dapat membuat prediksi yang lebih tepat, terutama dalam kasus di mana berita hoaks memiliki karakteristik bahasa yang serupa dengan berita faktual.
3. Menyeimbangkan *dataset* dengan mengumpulkan lebih banyak artikel hoaks yang memiliki panjang yang sebanding dengan artikel faktual. Hal ini akan membantu model dalam mempelajari pola-pola yang lebih representatif dan mengurangi bias terhadap panjang artikel.
4. Meningkatkan ketahanan model terhadap input teks yang pendek. Mengingat model saat ini cenderung mengklasifikasikan teks pendek sebagai hoaks, diperlukan strategi untuk mengatasi bias ini. Misalnya, dengan melatih model secara khusus pada *dataset* yang seimbang dalam hal panjang teks, atau dengan menggunakan teknik *data augmentation* untuk memperkaya variasi panjang teks dalam *dataset*.
5. Mengeksplorasi fitur-fitur tambahan selain teks, seperti informasi *meta* atau konteks, yang dapat memberikan sinyal tambahan untuk deteksi hoaks. Dengan memanfaatkan informasi yang lebih kaya, model dapat membuat prediksi yang lebih akurat dan mengurangi ketergantungan pada karakteristik bahasa semata.

Dengan mengatasi keterbatasan-keterbatasan yang ada dan menerapkan saran-saran tersebut, penelitian selanjutnya dapat mengembangkan model deteksi hoaks yang lebih robust dan andal. Model yang mampu mempertimbangkan akurasi faktual, menangani variasi panjang teks, dan memanfaatkan informasi tambahan akan lebih siap untuk menghadapi tantangan deteksi hoaks di dunia nyata.

U N I V E R S I T A S
M U L T I M E D I A
N U S A N T A R A