BAB V
KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang dilakukan dengan judul "Penggunaan Topic Modeling dengan Metode *latent dirichlet allocation* terhadap Berita Saham Indeks LQ45," terdapat sejumlah kesimpulan yang berhasil diperoleh seperti berikut ini:

1. Implementasi *topic modeling* untuk sekumpulan berita dengan metode *latent dirichlet allocation* dilakukan melalui beberapa tahapan proses. Tahap pertama, lakukan pra-pemrosesan data terhadap data yang terkumpul dengan *case folding, cleasing, tokenizing, stopwords removal*, dan *stemming* sebelum dilakukan pengimplementasian model sehingga data berbentuk *list of tokens*. Kedua, bagi dataset menjadi dua untuk data latih dan data uji dengan rasio 60:40. Ketiga, representasikan data latih ke dalam bentuk *bag of words* sebelum dilakukan pengimplementasian terhadap metode. Keempat, lakukan pengujian terhadap model dengan metrik *coherence score* mencari jumlah topik optimal, yaitu nilai *coherence score* tertinggi. Kelima, analisa hasil akhir berupa daftar topik dicetak berdasarkan jumlah topik dengan nilai *coherence score* tertinggi.
2. Berdasarkan analisis menggunakan metode *latent dirichlet allocation* (LDA), evaluasi *coherence score* dalam rentang 2 hingga 10 topik menunjukkan bahwa jumlah topik optimal adalah empat topik, dengan nilai tertinggi sebesar 0.604946 dan nilai terendah sebesar 0.390450 pada dua topik. Pada jumlah topik optimal, LDA menghasilkan empat topik utama dengan dengan tema 'Perbankan dan pertumbuhan kredit', 'Analisis teknikal dan pergerakan harga saham', 'Kinerja sektor saham dan pergerakan pasar', dan 'Saham perbankan dan pergerakan emiten besar'. Hasil ini menunjukkan bahwa LDA mampu mengidentifikasi pola-pola utama dalam data berita saham dengan hasil yang relevan dan bermakna.

3. Berdasarkan hasil perbandingan *coherence score*, metode LDA menunjukkan performa yang lebih tinggi dibandingkan dengan HDP dan NMF. Hal ini didasari oleh analisis statistika deskriptif metode LDA menunjukkan keunggulan dibandingkan HDP dan NMF dengan rata-rata tertinggi (0.559308) dan median terbesar (0.581115), yang menandakan kualitas topik yang lebih baik dan lebih konsisten. Meskipun memiliki standar deviasi yang lebih besar dibanding HDP, LDA tetap menghasilkan topik dengan keterkaitan kata yang lebih koheren. Sementara HDP lebih stabil dengan standar deviasi terkecil (0.019430), dan NMF memiliki fluktuasi terbesar serta performa terendah, hasil ini menunjukkan bahwa LDA adalah metode yang paling efektif dalam menghasilkan topik dengan kualitas terbaik.

5.2 Saran

Adapun saran yang dapat dilakukan berdasarkan penelitian ini untuk mengembangkan penelitian serupa:

1. Dalam hal sumber data, disarankan untuk memperluas cakupan dataset dengan menggabungkan berita saham dari berbagai platform seperti CNN, Kompas, atau portal berita lainnya yang memiliki cakupan lebih luas.
2. Menambahkan fitur dataset yang mencakup periode waktu secara *real-time* dan ter-*update* secara otomatis, sehingga pengguna dapat memperoleh topik yang lebih sesuai dengan dinamika dan kondisi pasar saham terkini.
3. Mengimplementasikan analisis tren topik berdasarkan rentang waktu tertentu untuk memahami bagaimana topik tertentu berkembang atau berubah sesuai dengan peristiwa pasar. Dengan memanfaatkan teknik analisis temporal, pengguna dapat mengidentifikasi pola-pola topik yang muncul sebelum atau setelah peristiwa ekonomi penting, sehingga memberikan wawasan yang lebih mendalam tentang hubungan antara berita saham dan pergerakan pasar.

4. Pada tahap pra-pemrosesan data, disarankan untuk menerapkan teknik n-grams, khususnya bigrams dan trigrams, guna menangkap istilah *multi*-kata yang sering muncul dalam berita saham, seperti “indeks LQ45” atau "harga saham". Pendekatan ini dapat meningkatkan pemahaman model terhadap konteks dan hubungan antar kata dalam teks.
5. Dalam hal representasi data, disarankan untuk menggunakan metode yang lebih kaya dan kompleks, seperti TF-IDF atau *word embeddings* (misalnya Word2Vec atau BERT), sebagai alternatif atau pelengkap dari *Bag of Words* (BoW). Representasi ini mampu menangkap hubungan semantik antar kata, sehingga dapat meningkatkan kualitas topik yang dihasilkan oleh model. Dengan demikian, kelemahan BoW yang hanya mempertimbangkan frekuensi kata tanpa memperhatikan konteks dapat diatasi.

Halaman ini sengaja dikosongi