

ABSTRACT

TIME SERIES ANALYSIS USING CONVOLUTIONAL NEURAL NETWORKS (CNN) AND RECURRENT NEURAL NETWORKS (RNN) FOR MODELING THE FREQUENCY OF INFECTIOUS DISEASE EPIDEMIC NEWS

By

Clarisa Septia Damayanti

The COVID-19 pandemic has highlighted the importance of leveraging online data as a tool for predicting future infectious disease trends. This study aims to compare the performance of two deep learning methods, namely Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), in predicting the daily frequency of online news publications based on three sentiment classes: negative, neutral, and positive. The results show that CNN delivers the best performance, with an RMSE of 0.14 and MAPE of 27%, demonstrating its superiority in recognizing complex patterns in large datasets, especially for negative and neutral sentiment data. Meanwhile, RNN also yields reasonably good results, particularly for smaller datasets such as those with positive sentiment, although with slightly lower accuracy (RMSE of 0.17 and MAPE of 35%). These findings suggest that CNN is highly recommended for predictions on large-scale datasets, while RNN serves as a relevant alternative when data availability is limited, albeit with a slightly lower accuracy rate. Overall, deep learning models have proven effective in predicting the frequency of online news publications based on sentiment, supporting the use of online news as an alternative data source for monitoring public health issues.

Keywords: COVID-19, infectious diseases, online news, sentiment.

ABSTRACT

TIME SERIES ANALYSIS USING CONVOLUTIONAL NEURAL NETWORKS (CNN) AND RECURRENT NEURAL NETWORKS (RNN) FOR MODELING THE FREQUENCY OF INFECTIOUS DISEASE EPIDEMIC NEWS

By

Clarisa Septia Damayanti

Pandemi COVID-19 telah menyoroti pentingnya pemanfaatan data online sebagai alat untuk memprediksi tren penyakit menular di masa depan. Studi ini bertujuan untuk membandingkan kinerja dua metode deep learning, yaitu Convolutional Neural Network (CNN) dan Recurrent Neural Network (RNN), dalam memprediksi frekuensi harian publikasi berita online berdasarkan tiga kelas sentimen: negatif, netral, dan positif. Hasil penelitian menunjukkan bahwa CNN memberikan kinerja terbaik, dengan nilai RMSE sebesar 0,14 dan MAPE sebesar 27%, menunjukkan keunggulannya dalam mengenali pola kompleks pada dataset besar, terutama untuk data sentimen negatif dan netral. Sementara itu, RNN juga menghasilkan performa yang cukup baik, khususnya untuk dataset yang lebih kecil seperti data dengan sentimen positif, meskipun dengan tingkat akurasi yang sedikit lebih rendah (RMSE sebesar 0,17 dan MAPE sebesar 35%). Temuan ini menunjukkan bahwa CNN sangat direkomendasikan untuk prediksi pada dataset berskala besar, sementara RNN merupakan alternatif yang relevan ketika ketersediaan data terbatas, meskipun dengan tingkat akurasi yang sedikit lebih rendah. Secara keseluruhan, model deep learning terbukti efektif dalam memprediksi frekuensi publikasi berita online berdasarkan sentimen, sehingga mendukung penggunaan berita online sebagai sumber data alternatif untuk pemantauan isu kesehatan masyarakat.

Kata kunci: COVID-19, penyakit menular, berita online, sentimen.