Penggalian Informasi Untuk Identifikasi Permintaan Bantuan Korban Bencana Alam Menggunakan Data Twitter = Extracting Information to Identify Assistance for Natural Disaster Victims Using Twitter Data

Deden Ade Nurdeni, author

Abstrak

Kajian risiko bencana di Indonesia oleh BNPB menunjukkan jumlah jiwa terpapar risiko bencana tersebar di seluruh Indonesia dengan total potensi jiwa terpapar lebih dari 255 juta jiwa. Hasil kajian ini menunjukkan bahwa dampak bencana di Indonesia terbilang sangat tinggi. Sistem penanggulangan khususnya pada masa tanggap darurat menjadi hal yang krusial untuk dapat meminimalisir risiko. Namun, pemberian bantuan kepada korban bencana terkendala beberapa hal, antara lain keterlambatan dalam penyaluran, kurangnya informasi lokasi korban, dan distribusi bantuan yang tidak merata. Untuk memberikan informasi yang cepat dan tepat, BNPB telah membangun beberapa sistem informasi seperti DIBI, InAware, Geospasial, Petabencana.id dan InaRisk. Akan tetapi tidak secara realtime menampilkan wilayah terdampak bencana dengan memnunjukkan jenis kebutuhan bantuan apa yang dibutuhkan korban pada saat itu. Untuk memberikan solusi atas permasalah tersebut, penelitian ini membangun model yang mampu mengklasifikasikan data teks dari Twitter terkait bencana kedalam jenis bantuan yang diminta oleh korban bencana secara realtime. Selain itu visualisasi berupa dashboard dibangun dalam bentuk aplikasi berbasis peta untuk menampilkan lokasi korban yang terdampak. Penelitian ini mengunakan teknik text mining mengolah data Twitter dengan pendekatan metode klasifikasi multi label dan ekstraksi informasi lokasi menggunakan metode Stanford NER. Algoritme yang digunakan adalan Naive Bayes, Support Vector Machine, dan Logistic Regression dengan kombinasi metode tranformasi data multi label OneVsRest, Binary Relevance, Label Power-set, dan Classifier Chain. Representasi teks menggunakan N-Grams dengan pembobotan TF-IDF. Model terbaik untuk klasifikasi multi label pada penelitian ini adalah kombinasi Support Vector Machine dan Clasifier Chain dengan fitur UniGram+BiGram dengan nilai precision 82%, recall 70%, dan F1-score 75%. Stanford NER menghasilkan F1-score 83% untuk klasifikasi lokasi yang menjadi masukan untuk teknik geocoding. Hasil geocoding berupa informasi spasial ditampilkan dalam bentuk dashboard berbasis peta.

..... The study of disaster risk in Indonesia by BNPB shows the number of people exposed to disaster risk throughout Indonesia with a total potential life of 255 million people. The results of this study indicate that the impact of disasters in Indonesia is quite high. The response system, especially during the emergency response period, is crucial to be able to minimize risks. However, providing assistance to disaster victims is hampered by several things, including delays in providing assistance, lack of information on the location of victims, and uneven distribution of aid. To provide fast and accurate information, BNPB has built several information systems such as DIBI, InAware, Geospatial, Petabencana.id and InaRisk. However, it does not display the disaster area in real-time by showing what kind of assistance needs the victim needs at that time. To provide a solution to these problems, this study builds a model that is able to classify text data from Twitter related to the type of assistance requested by disaster victims in real-time. In addition, a dashboard is built in the form of a map-based application to display the location of the realized victim. This study uses text mining techniques to process Twitter data with a multi-label classification approach and location

information extraction using the Stanford NER method. The algorithms used are Naive Bayes, Support Vector Machine, and Logistic Regression with a combination of OneVsRest, Binary Relevance, Power-set Label, and Classifier Chain. Text representation using N-Grams with TF-IDF weighting. The best model for multi-label classification in this study is a combination of Support Vector Machine and Classifier Chain with UniGram+BiGram features with 82% precision, 70% recall, and 75% F1-score. Stanford NER produces an F1-score of 83% for location classification which is the input for geocoding techniques. Geocoding results in the form of spatial information are displayed in a map-based dashboard.