

Klasifikasi ketinggian banjir menggunakan media sosial Twitter berbahasa Indonesia = Classification of inundation level using Indonesian Twitter

Kwee Felicia Ilona, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920517805&lokasi=lokal>

Abstrak

Kejadian banjir ekstrem diperkirakan semakin sering terjadi seiring dengan perubahan iklim yang belum menunjukkan tanda-tanda perbaikan. Hal ini berpotensi menyebabkan curah hujan yang lebih tinggi dari rata-rata dan laju kenaikan genangan banjir yang lebih cepat dari biasanya. Ketidaktahuan bahwa suatu daerah tergenang banjir juga bisa mengakibatkan kendaraan terjebak di daerah banjir, kemacetan lalu lintas, serta terlambatnya evakuasi warga terutama yang tinggal di daerah rawan banjir. Penelitian ini bertujuan memberikan alternatif sumber informasi mengenai ketinggian genangan banjir dengan memanfaatkan data teks dari tweet pada media sosial Twitter. Salah satu tantangannya yaitu bahwa ketinggian genangan tidak selalu disebutkan dalam standar satuan internasional seperti centimeter atau meter sehingga machine learning digunakan untuk mengatasinya. Penyebutan ketinggian genangan didapati bisa menggunakan referensi bagian tubuh seperti lutut dan pinggang, serta juga bagian kendaraan atau kondisi jalan. Model yang diusulkan memberikan dua keluaran, yaitu kategori relevansi tweet terhadap informasi ketinggian banjir (Relevan atau Tidak Relevan) dan kategori ketinggian banjir (Tinggi, Sedang, Rendah, dan Tidak Diketahui). Algoritma klasifikasi yang digunakan yaitu SVM (Linear SVC dan RBF), Logistic Regression, Random Forest, Decision Tree, dan Naïve Bayes. Hasil uji coba menunjukkan bahwa nilai akurasi tertinggi untuk klasifikasi relevansi tweet adalah 91% dan F1-score tertinggi sebesar 82% diperoleh dengan menggunakan algoritma SVM Linear SVC. Sedangkan hasil klasifikasi ketinggian genangan terbaik diperoleh saat menggunakan SVM Linear SVC dengan akurasi 83% dan rata-rata F1-score 70%.

.....Extreme flood events are expected to occur more frequently as climate change has yet to show signs of improvement. This has the potential to lead to higher rainfall and floods that come more quickly. This has the potential for vehicle trapping, traffic jams, or delay in evacuation for people who live in areas which are prone to flooding. Hence, this study aims to provide an alternative source of information in flood conditions by using data in social media Twitter. One of the challenges was information about inundation level is not always in international standard unit like centimeter or meter so that machine learning was used to cope with this problem. Mention of inundation level was found to be done by also referring to certain body parts like knee and waist, and also parts of vehicles or road condition. The proposed model is expected to provide two outputs, which are relevance category of tweet (Relevant or Irrelevant) and inundation level category (High, Medium, Low, Unknown). Some classifier algorithms were used, like SVM (Linear SVC and RBF), Logistic Regression, Random Forest, Decision Tree, and Naïve Bayes. The test results showed that the best relevance classification resulted in 91% accuracy (SVM Linear SVC) and 82% average F1-score by using SVM Linear SVC. On the other side, the best result of classification of inundation level was obtained when using SVM Linear SVC which resulted in 83% accuracy and 70% average F1- score.