

Deskripsi Umum :

**Model Klasifikasi Data Bencana Alam Dari Situs Berita Dengan
Metode Support Vector Machine**

Abstract - Natural disasters are accident caused by nature such as floods, earthquakes, and forest fires. This raises concerns and has a negative impact on society. The development of increasingly sophisticated technology accelerates the dissemination of information, one of which is news sites. One of the news site's topics contained is natural disasters. The abundance of news information makes it difficult for the public to sort data by type of natural disaster. Therefore, a system was built that aims to classify the types of natural disasters such as floods, earthquakes, and forest fires. The data used from several news sites include Kompas.com and liputan6.com. The data is collected using crawling technique, then the data is further processed through the preprocessing stage to produce structured data. The data is classified using the Support Vector Machine algorithm. The accuracy value obtained from the classification is 83.33%.

Keywords – natural disaster, crawling, classification, news site, vector machine

Abstrak - Bencana alam merupakan peristiwa yang diakibatkan oleh alam seperti banjir, gempa, dan kebakaran hutan. Hal ini menimbulkan kekhawatiran dan dampak negatif yang merugikan bagi masyarakat. Perkembangan teknologi yang semakin canggih mempercepat dalam penyebaran informasi salah satunya situs berita. Topik yang terdapat dalam situs berita salah satunya adalah bencana alam. Banyaknya informasi berita membuat masyarakat sulit untuk memilah data berdasarkan jenis bencana alam. Oleh karena itu dibangun sebuah sistem yang bertujuan untuk mengklasifikasikan jenis bencana alam banjir, gempa, dan kebakaran hutan. Data yang digunakan berasal dari beberapa situs berita diantaranya kompas.com dan liputan6.com. Data tersebut dihimpun menggunakan teknik *crawling*, kemudian selanjutnya diolah melalui tahap *preprocessing* untuk menghasilkan data yang terstruktur. Data tersebut di klasifikasi menggunakan algoritma *Support Vector Machine*. Nilai akurasi yang didapat dari klasifikasi adalah sebesar 96%.

Kata Kunci – bencana alam, *crawling*, klasifikasi, situs berita, *support vector machine*

Pendahuluan

Bencana alam merupakan bencana yang diakibatkan oleh peristiwa yang disebabkan oleh alam seperti banjir, gempa, dan kebakaran hutan. Bencana alam seringkali mengakibatkan dampak negatif bagi masyarakat baik itu kerusakan, kerugian materil, dan korban jiwa. Indonesia rawan terhadap bencana alam dikarenakan terletak pada pertemuan tiga lempeng yaitu lempeng Eurasia, lempeng Pasifik, dan lempeng Australia.

Berdasarkan data Badan Nasional Penanggulangan Bencana (BNPB) sepanjang bulan Agustus 2021, telah terjadi sebanyak 155 kejadian bencana alam di Indonesia, dan dampak dari bencana alam tercatat korban meninggal 9 orang. Meskipun jika dilihat pada bulan yang sama antara 2020 dan 2021 jumlah kejadian lebih sedikit, tetapi dampak negatif yang ditimbulkan ternyata lebih besar.

Perkembangan teknologi yang semakin canggih menyebabkan penyebaran informasi juga semakin cepat, salah satu media penyebaran informasi adalah melalui situs berita. Situs berita adalah sebuah situs yang menyajikan berbagai informasi salah satunya adalah bencana alam. Banyaknya informasi yang terdapat dalam situs berita membuat masyarakat harus memilah informasi apakah berita tersebut termasuk dalam berita bencana alam atau non-bencana alam. Dalam mengklasifikasikan data berita secara manual menjadi berita bencana dan non-bencana adalah hal yang sulit dan memakan waktu yang cukup lama

Banyaknya informasi yang terdapat dalam situs berita bencana alam membuat masyarakat harus memilah informasi apakah berita tersebut termasuk dalam berita bencana alam banjir, gempa, atau kebakaran hutan. Klasifikasi data berita secara manual adalah hal yang sulit dan kurang efektif karena membutuhkan waktu yang lama dalam melakukannya (Fernandes *et al.*, 2021).

Pada penelitian (Domala *et al.*, 2020) dilakukan klasifikasi terhadap data pada situs-situs berita, berita diklasifikasikan menjadi berita bencana yang relevan atau berita bencana yang tidak relevan. Metode klasifikasi yang digunakan pada penelitian ini yaitu *Naive Bayes*, *logistic regression*, *SVM*, *xtreme gradient boosting* dan *random forest*. Beberapa situs berita yang digunakan pada penelitian ini diantaranya Times of India, NDTV India dan India express. Hasil dari penelitian ini menunjukkan nilai akurasi terbesar yaitu 89% dalam pendeteksian berita tentang bencana alam dengan menggunakan algoritma *logistic regression*.

Pada penelitian (Gopal *et al.*, 2020) dilakukan klasifikasi apakah suatu berita dikatakan relevan atau tidak relevan mengenai bencana alam. Penelitian ini membangun *crawler* untuk mendapatkan data yang berasal dari berbagai situs berita. Klasifikasi yang digunakan

menggunakan beberapa algoritma diantaranya *K-nearest Neighbor*, *Support Vector Machines*, *Multi Layer Perceptrons*, dan *Logistic Regression*. Algoritma klasifikasi digunakan untuk memprediksi relevan atau tidak relevan. Pada Algoritma-yang digunakan dilakukan komparasi menggunakan unigram, bigram, dan trigram. Hasil dari penelitian ini adalah bigram menghasilkan hasil yang lebih baik dibandingkan dengan unigram dan trigram. Pada hasil rerata, Support Vector Machine menghasilkan rata-rata yang lebih baik daripada algoritma lainnya.

Sistem yang dibangun pada penelitian ini merupakan sistem klasifikasi berdasarkan data yang berasal dari situs berita menggunakan metode *Support Vector Machine* (SVM). Sumber data berita yang digunakan diantaranya *kompas.com* dan *liputan6.com*. Data tersebut dihimpun menggunakan teknik *crawling*. Data yang telah dihimpun diolah menggunakan *preprocessing* agar menjadi data yang terstruktur. Tahap selanjutnya adalah dilakukan pembobotan kata menggunakan TF-IDF. Kemudian data dibagi menjadi dua bagian yaitu data *training* dan data *testing*. Selanjutnya data tersebut melalui serangkaian proses klasifikasi hingga menghasilkan akurasi. Nilai akurasi yang didapatkan dengan algoritma *Support Vector Machine* adalah sebesar 96%.

RERERENCES

- Domala, J. *et al.* (2020) ‘Automated Identification of Disaster News for Crisis Management using Machine Learning and Natural Language Processing’, *Proceedings of the International Conference on Electronics and Sustainable Communication Systems, ICESC 2020*, (Icesc), pp. 503–508. doi: 10.1109/ICESC48915.2020.9156031.
- Fernandes, C. *et al.* (2021) ‘Automated Disaster News Collection Classification and Geoparsing’, *SSRN Electronic Journal*, (Icsmdi). doi: 10.2139/ssrn.3852688.
- Gopal, L. S. *et al.* (2020) ‘Machine Learning based Classification of Online News Data for Disaster Management’, *2020 IEEE Global Humanitarian Technology Conference, GHTC 2020*. doi: 10.1109/GHTC46280.2020.9342921.