

ABSTRAK

Media sosial yang paling terkenal dan paling banyak dipakai masyarakat umum adalah Twitter dan media sosial ini memiliki pengguna dalam jumlah besar di seluruh belahan dunia dan digunakan baik sebagai media komunikasi maupun sebagai media untuk membagikan informasi di internet. Tujuan penelitian ini adalah untuk mengklasifikasikan data yang telah dikumpulkan sebelumnya dari Twitter dan tidak hanya untuk mengklasifikasikan data tersebut, tapi juga penelitian ini berfungsi untuk membandingkan dua metode algoritma data mining yang paling sering digunakan dalam metode klasifikasi. Kedua algoritma ini adalah Naive Bayes Algorithm dan K-Nearest Algorithm dalam algoritma data mining yang digunakan dalam pengklasifikasian data. Terdapat perbedaan hasil klasifikasi yang jauh pada kedua algoritma baik di performa maupun akurasi, yang merupakan tujuan utama penelitian ini. Untuk membuktikan dan membandingkan performa dan akurasi dari hasil klasifikasi, penelitian ini perlu dilakukan secara sistematis. Langkah-langkah untuk mengumpulkan data-data itu sendiri, membersihkan data sehingga menjadi data yang layak diproses, memproses data melalui algoritma data mining, memvisualisasikan hasil dari proses, dan yang terakhir menampilkan hasilnya dan membandingkannya. Dengan perangkat data mining berupa R Studio sebagai alat untuk mengumpulkan data, menciptakan dataset untuk diproses dalam penelitian, memproses dataset dengan algoritma klasifikasi, dan menganalisa performa dengan memvisualisasikan hasilnya dengan menggunakan metode confusion matrix. Di tahapan terakhir, kandungan dalam confusion matrix akan digunakan sebagai nilai yang dibutuhkan untuk menghitung skor F1 dan membandingkannya dengan hasil secara keseluruhan dari kedua algoritma klasifikasi, yang akan diketahui algoritma mana yang menghasilkan skor yang lebih besar dan lebih cocok dalam memproses dataset dalam penelitian ini. Dari penelitian ini, didapatkan hasil bahwa algoritma Naive Bayes lebih baik dalam klasifikasi dan membersihkan dataset dibandingkan algoritma K-Nearest Neighbor baik dalam akurasi dan performa. Meskipun keakuratan kedua algoritma memiliki perbedaan yang tipis, terdapat perbedaan performa yang besar diantara Naive Bayes dan K-Nearest Neighbor. Dengan hasil dari

penelitian ini, juga membantu mengembangkan algoritma klasifikasi data dalam penelitian selanjutnya untuk referensi masa yang akan datang.

Kata kunci—*Covid-19, Klasifikasi, Algoritma, Naïve Bayes, K-Nearest Neighbor, Confusion Matrix, F1 Score.*