

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Mengacu pada tahapan proses yang dijelaskan pada bab sebelumnya, akan dipeloreh beberapa simpulan sebagai berikut:

1. Pada penelitian pengambilan data Twitter dengan memanfaatkan tweet dari pengguna Twitter berhasil dilakukan menggunakan metode *crawling* dengan bantuan API dan pengambilan data tweet menggunakan kata kunci “vaksin covid” dari Twitter dan berhasil di dapatkan sebanyak 109.481 tweet tetapi dengan banyaknya tweet terduplikat dan tweet tidak berbahasa ndonesia maka dataset yang bisa digunakan hanya sebanyak 8000 tweets.
2. Dalam penelitian ini untuk sentimen analisis vaksin covid menggunakan algoritma *naïve bayes classifier* terbukti akurat karena menghasilkan inilai akurasi sebesar 73,1% , inilai *precision* sebesar 73%, inilai *recall* sebesar 97%, dan inilai *f1-score* sebanyak 83%
3. Visualisasi data dari hasil klasifikasi terdapat 3 *word cloud*.

- a. *Word cloud* keseluruhan dataset

Pada *word cloud* keseluruhan untuk melihat kecenderungan sentimen dikarenakan dataset untuk kelas positif lebih banyak di banding dataset kelas negatif dan didapatkan hasilnya *word cloud* keseluruhan cenderung pada sentimen positif. Pada *word cloud* keseluruhan dataset “vaksin covid” menjadi topik yang sering dibicarakan. Selain tu terdapat juga kata “uji klinis”, “dosis vaksin”, “terima vaksin”.

- b. *Word cloud* positif

Pada *word cloud* positif sama seperti *word cloud* keseluruhan dataset “vaksin covid” menjadi topik yang sering di bicarakan lalu terdapat juga kata “uji klinis”, “sedia vaksin” dan “gratis vaksin”. Pada *word cloud* positif masyarakat lebih beropini seperti pendistribusian vaksin, berapa harga vaksin, apakah vaksin gratis, dan juga kandidat yang mendapatkan vaksin

c. *Word cloud* negatif

Pada *word cloud* negatif sama seperti *word cloud* keseluruhan dataset dan *word cloud* positif “vaksin covid” masih menjadi topik yang sering dibicarakan dan terdapat juga kata seperti “tular covid”, “takut”, dan “efek samping”. Pada *word cloud* negatif masyarakat lebih beropini seperti covid menular, efek samping dari vaksin, dan masyarakat takut terhadap vaksinasi.

5.2 Saran

Berikut adalah beberapa rekomendasi berdasarkan temuan penelitian:

1. Dalam penelitian ini menggunakan data yang di dapatkan dari hasil metode *crawling* menggunakan Bahasa program python tetapi tetapi program yang penulis buat masih belum benar sehingga dataset vaksin covid yang didapatkan tidak terlalu bagus dikarenakan tweet yang berhasil masuk ke program tidak full seperti atau terpotong seperti “Kehalalan vaksin\n\nKehalalan vaksin merupakan hal sensitif. Calon vaksin Covid-19 saat ini belum dipastikan kehalala\xe2\x80\xa6 <https://t.co/UmH56XNQpN>” dan jika masuk ke tahap preprocessing akan seperti “kehalalan vaksin\nkehalalan vaksin merupakan hal sensitif calon vaksin covid saat ini belum dipastikan kehalalaxexa”.
2. Dalam penelitian ini labeling menggunakan metode *lexicon based* secara otomatis, yaitu dengan menghitung setiap kata apakah terdapat di dalam dataset lexicon tetapi 84% kata di data vaksin covid tidak terdapat di dalam dataset lexicon dan hanya 16% yang bisa dilabel kata positif atau negatif. Untuk penelitian selanjutnya, dapat menggunakan dataset yang lebih lengkap dan di labelkan oleh ahli Bahasa
3. Hanya satu set data yang digunakan dalam penelitian ini yaitu data tweet vaksin covid dalam bahasa Indonesia. Untuk penelitian selanjutnya, data dalam bahasa asing seperti Bahasa Inggris dapat ditambahkan.

4. Dalam penelitian ini dataset positif dan negatif tidak seimbang, lebih banyak dataset positif yang membuat hasil lebih besar ke positif. Untuk penelitian selanjutnya, dapat menggunakan dataset yang seimbang kelas positif dan negatifnya
5. Dalam penelitian ini hanya menggunakan satu algoritma yaitu *naïve bayes classifier*. Untuk penelitian selanjutnya, dapat menggunakan perbandingan algoritma sebagai pembanding hasil uji model, seperti algoritma support vector machine (SVM)

