

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan penelitian dengan judul Perbandingan Metode *Random Forest* dan *K-Nearest Neighbors* pada Analisis Sentimen Pengguna Twitter mengenai Promo Gojek yang telah dilakukan, dapat disimpulkan dalam poin berikut ini

1. Penelitian ini menggunakan data yang diperoleh dari media sosial Twitter melalui proses *crawling* data dengan kurun waktu pengumpulan data dari 01 April 2022 hingga 30 Juni 2022. Data yang berhasil dikumpulkan sebanyak 453 tweet berupa opini publik tentang ketersediaan promo Gojek pada perangkat pengguna, lalu data tersebut diberi label positif dan negatif oleh tiga penilai secara manual yang menghasilkan 209 data positif, dan 244 data negatif. Tahap selanjutnya adalah tahap *preproses data*, data yang telah diberi label akan melewati beberapa tahap yang diantaranya, *case folding*, *cleansing data*, normalisasi, *stemming*, *stopword removal*, dan tokenisasi. Setelah melalui *preproses*, data akan diberi bobot menggunakan metode TF-IDF sebelum masuk proses klasifikasi. Setelah itu, data akan melalui proses klasifikasi menggunakan metode *K-Nearest Neighbors* dan menggunakan *10-Fold Cross Validation*, yang dimana data akan dibagi secara acak dan adil menjadi 80% data latih (*training*) dan 20% data uji (*testing*). Setelah diklasifikasi, diperoleh nilai K terbaik yang terdapat pada $K = 7$ yang menghasilkan performa berupa nilai akurasi sebesar 75%, nilai presisi 80%, nilai *recall* 64%, dan nilai *f-1 score* sebesar 71%.
2. Pengujian kedua dilakukan dengan menggunakan metode klasifikasi *Random Forest*, dengan menggunakan data yang sama yaitu data berupa opini publik tentang ketersediaan promo Gojek pada perangkat pengguna sebanyak 453 tweet. Sebelum dilakukan klasifikasi menggunakan *Random Forest*, data diolah seperti sebelumnya, yaitu melakukan pelabelan data secara manual, *preprocess data*, pembobotan kata menggunakan TF-IDF. Lalu pada tahap *Random Forest* terdapat

parameter yang penguji lakukan diantaranya adalah *n estimators* [100, 200, 300, 500, 1000], *max features* [auto, sqrt, log2], *max depth* [3, 4, 5, 6, 7, 8, 80, 90, 100, 110], dan *criterion* [gini, entropy]. Berdasarkan percobaan yang dilakukan, hasil terbaik diperoleh pada parameter *criterion* = entropy, *max depth* = 100, *max features* = auto, dan *n estimators* = 1000, yang menghasilkan *mean test* terbaik dengan nilai 0.75198863636363. Lalu menghasilkan nilai akurasi sebesar 77%, nilai presisi sebesar 75%, nilai *recall* sebesar 77%, dan nilai *f-1 score* 75%.

3. Dari hasil evaluasi kedua metode klasifikasi diatas, dapat dilakukan perbandingan. Metode *Random Forest* menghasilkan nilai evaluasi yang lebih baik daripada metode *K-Nearest Neighbors* untuk mengolah data promo gojek pada penelitian ini.
4. Pada penelitian ini, hasil evaluasi menunjukkan hasil dibawah 80%, dimana hasil tersebut masih kurang memuaskan, dikarenakan pada saat penelitian ini berlangsung, ketika proses crawling data, data yang diambil adalah sebanyak 2000 data, tetapi dari 2000 data tersebut masih banyak data yang duplikat, sehingga data yang bisa digunakan hanya sebanyak 453 data. Pada saat melakukan preprocess juga masih banyak noise yang masih masuk kedalam proses klasifikasi sehingga, hal tersebut berdampak pada proses klasifikasi, dikarenakan metode *Random Forest* dan *K-Nearest Neighbors* adalah metode klasifikasi yang cocok untuk mengklasifikasi data berjumlah besar, sedangkan pada penelitian ini hanya 453 data saja yang digunakan.

5.2 Saran

Berikut adalah beberapa saran yang dapat diberikan dan dipertimbangkan untuk penelitian di masa mendatang.

1. Jumlah data tweet yang digunakan dapat diperbanyak serta rentang waktu pengambilan data bisa lebih diperpanjang, sehingga performa dari pemodelan klasifikasi dengan metode *K-Nearest Neighbors* dan *Random Forest* mendapatkan hasil yang lebih memuaskan, dikarenakan metode tersebut cocok untuk pemodelan klasifikasi yang memiliki data berkuantitas tinggi.

2. Pada *preprocess*, diharapkan lebih diperhatikan dan diteliti secara merinci lagi, seperti lebih diperhatikan lagi pada saat memasukkan data *stopword* dan normalisasi, karena dilakukan secara manual, maka lebih baik untuk diteliti secara merinci lagi. Sehingga pada saat proses klasifikasi, hasil bisa menunjukkan hasil yang lebih tinggi.
3. Diharapkan pada penelitian di masa mendatang, dapat melakukan analisis dengan menggunakan metode klasifikasi lain seperti *Naïve Bayes* atau *Support Vector Machine*. Karena kedua metode tersebut cocok untuk melakukan pemodelan klasifikasi dengan data yang berkuantitas rendah.