

BAB I

PENDAHULUAN

1.1 Latar Belakang

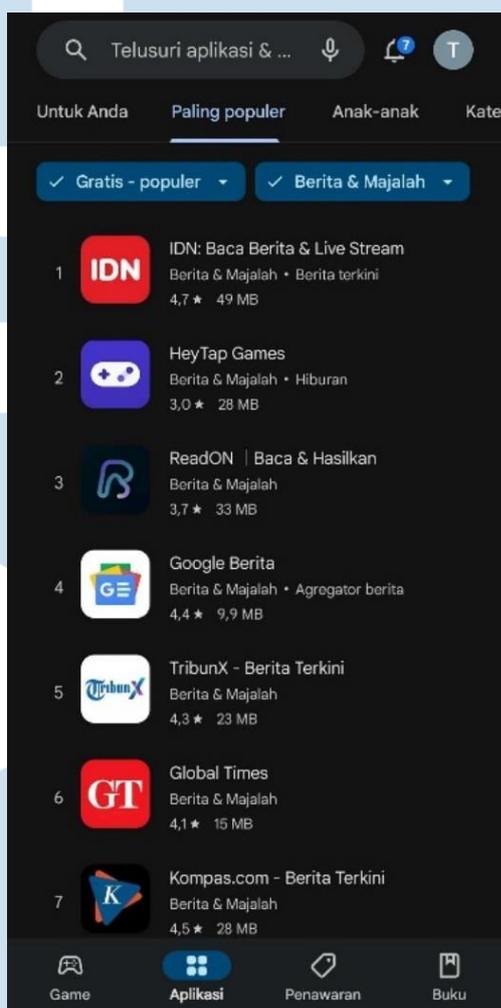
Google Play store merupakan sebuah aplikasi *multiplatform* yang dapat diakses melalui *website* atau *mobile* [1]. Fungsi dari *Google playstore* menyediakan layanan untuk memudahkan pengguna menemukan dan mengunduh produk yang dicari. Produk yang di tawarkan dalam *google playstore* sangat beragam mulai dari aplikasi, film, musik, atau buku. Perpaduan dengan perkembangan penyebaran informasi dengan perkembangan internet hingga adanya *Google Play store* menghasilkan. Salah satu aplikasi yang terdapat pada *Google Play store* yaitu *IDN News*. *IDN News* merupakan sebuah aplikasi hiburan yang dapat memberikan berbagai macam hiburan seperti kuis, *live stream*, dan fokus utama dari aplikasi yaitu berita.



Gambar 1. 1 Aplikasi IDN pada google playstore.

Aplikasi *IDN news* merupakan aplikasi yang dimiliki oleh *IDN media*. Aplikasi *IDN* telah dirilis sejak 8 juni 2020. Berawal sebagai aplikasi news agregator yang menampilkan berita dari berbagai 4 *platform* media digital. 4 media digital tersebut yaitu *IDN News*, *Popbela.com*, *Popmama.com*, dan *Duniaku.com*. konten berita yang terdapat pada aplikasi sangat beragam mulai dari sepak bola, aplikasi *IDN* terus mengalami perkembangan hingga pada tahun 2021 *IDN News* merilis fitur *live stream* dengan nama *IDN Live*. Fitur *live* tidak hanya diisi oleh konten berita

saja. Terdapat *live gaming* dengan berkerja sama dengan berbagai tim *e-sport* seperti Bigetron, Aura *Esport*, konten hiburan dengan JKT48, *UP Girls*, *Typicals*, serta terdapat konten *live stream* horor. Aplikasi IDN news telah menjadi aplikasi paling populer nomor satu pada kategori “Berita & Majalah” pada *Google Play Store*.



Gambar 1. 2 Aplikasi Berita paling populer

Aplikasi IDN News memiliki beberapa fitur utama seperti berita, kuis, *live stream*. Aplikasi IDN telah di *download* sebanyak lebih dari 1 juta kali pada *Play Store*. serta telah mendapatkan ulasan/*feedback* dengan *rating* 4.7. *Feedback* yang diberikan oleh pengguna tidak selalu berisikan tentang sebuah hal yang positif namun ada juga *feedback* negatif hal. pengguna yang memberikan *feedback* negatif karena memiliki banyak keluhan terkait aplikasi IDN. Ulasan yang telah diberikan

dapat mempengaruhi keputusan calon pengguna aplikasi untuk mengunduh [2]. Ulasan yang di berikan para pengguna menjadi sangat penting bagi pengguna untuk memilih aplikasi, serta ulasan pengguna juga dijadikan sebagai informasi untuk *developer* untuk arah pengembangan aplikasinya. Peran sentimen analisis sangat dibutuhkan untuk mengetahui respon dari pengguna aplikasi. Pada penelitian ini dilakukan menggunakan ulasan aplikasi IDN. Pemilihan aplikasi IDN dikarenakan aplikasi IDN merupakan aplikasi yang paling populer nomor satu pada *Google Playstore* dengan kategori “Berita & Majalah” [3].

Sentimen analisis adalah sebuah proses untuk dapat menganalisis opini seseorang terhadap sebuah produk, nantinya hasil analisis tersebut akan menghasilkan sebuah evaluasi/kesimpulan mengenai penilaian seseorang terhadap produk. Sentimen analisis merupakan pengolahan data untuk mengetahui evaluasi, penilaian, emosi tentang suatu produk [4]. Untuk melakukan sentimen analisis terdapat banyak sekali data yang bernama *big data*. secara mudah *big data* adalah sekumpulan data yang sangat besar yang di dalamnya membuat berbagai macam informasi [5].

Pada penelitian ini akan melakukan klasifikasi terhadap ulasan/*feedback* mengenai aplikasi IDN pada *Google Play store*. Klasifikasi aplikasi dilakukan untuk mengetahui sentimen dari para penggunanya positif atau negatif, juga mengetahui kata apa saja yang paling sering muncul dari opini pengguna [6]. Dalam penelitian ini akan dilakukan proses pelabelan dari ulasan dengan rating yang telah diberikan. proses pelabelan terbagi menjadi dua kelas. Kelas pertama adalah kelas positif dengan rating ulasan 1-3, kemudian kelas ke dua adalah kelas positif dengan rating 4-5. Tidak terdapat kelas netral dalam pelabelan ini dikarenakan sulitnya untuk mengidentifikasi opini netral. Terdapat banyak metode yang dapat digunakan dalam teknik sentimen analisis. Pada penelitian terdahulu [7] algoritma *Random Forest* mendapatkan hasil akurasi yang optimal dengan nilai 83%, sedangkan algoritma KNN mendapatkan hasil yang lebih baik lagi dengan nilai 89%. Pada penelitian [8] mendapatkan nilai akurasi yang tinggi menggunakan algoritma Decision Tree dengan hasil akurasi sebesar 99.89%. Pada penelitian terdahulu [9] hasil akurasi Random Forest mendapatkan hasil sebesar 74%, Decision Tree

mendapatkan hasil akurasi sebesar 63%, KNN mendapatkan hasil 40%. Hasil pada penelitian [9] mendapatkan hasil yang kurang baik dikarenakan pada penelitian tersebut tidak terdapat feature selection untuk meningkatkan kinerja algoritma.

Berdasarkan penelitian [6][7] dalam menggunakan algoritma untuk teknik analisis maka, dalam melakukan klasifikasi pada penelitian ini menggunakan tiga metode yaitu *Random Forest*, *Decision Tree*, dan *K-Nearest Neighbor*. *Random Forest* adalah algoritma yang dapat digunakan untuk melakukan klasifikasi data dengan jumlah yang banyak. Cara kerja metode *Random Forest* adalah dengan mengkombinasikan pohon yang berasal dari model *decision tree* sehingga mendapatkan model terbaik. Kelebihan algoritma *Random Forest* mampu menangani data dengan beragam fitur, kekurangannya memerlukan sumber daya komputasi yang besar [10]. *Decision Tree* merupakan salah satu metode klasifikasi yang digambarkan seperti bentuk pohon dimana pohon tersebut memiliki cabang untuk menunjukkan beberapa pilihan alternatif. Dalam *Decision Tree* terdapat tiga jenis *node*, yaitu *root node*, *internal node*, *leaf node*. Kelebihan algoritma *decision tree* mampu menangani data yang memiliki noise, tidak memerlukan normalisasi data. kekurangannya adalah rentan *overfitting* terutama jika pohon terlalu kompleks [11]. *K Nearest Neighbors* merupakan algoritma klasifikasi yang termasuk *supervised learning* dimana cara kerja algoritma ini dengan mengklasifikasi data berdasarkan *train dataset* kemudian mencari dan mengambil nilai “K” ke tetangga terdekat. Kelebihan KNN tangguh terhadap data yang sangat besar. Kekurangannya perlu menentukan nilai K, pada penentuan nilai k tidak memiliki parameter yang jelas jarak nilai K yang harus digunakan [12].

Penelitian terdahulu telah membuktikan bahwa ketiga algoritma yang akan digunakan untuk penelitian memiliki akurasi sangat optimal, rata-rata hasil akurasi diatas 80%. Pada penelitian ini akan dilakukan komparasi pada sentimen analisis IDN. Komparasi yang dilakukan adalah dengan membandingkan tingkat *Accuracy*, *Precision*, *Recall*, dan *F1-score*. Tujuan dari komparasi metode *Random Forest*, *Decision Tree*, Dan *K-Nearest Neighbors* adalah untuk membandingkan dan mengetahui metode mana yang memiliki kinerja paling baik. Perbandingan algoritma *Random Forest*, *Decision Tree* dilakukan karena terdapat kemiripan cara

kerja algoritma, kedua algoritma tersebut bekerja dengan pohon keputusan yang membedakan hanya Random Forest terdapat lebih dari satu pohon sedangkan Decision Tree hanya satu pohon. Perbandingan terhadap KNN dilakukan karena algoritma KNN memiliki cara kerja yang paling berbeda diantara dua algoritma lainnya. Pengukuran performa metode dilakukan pada penelitian sentimen analisis aplikasi IDN. Penilaian metode yang lebih baik dilakukan berdasarkan pengukuran nilai akurasi, *precision*, *recall*, dan *f1_score*.

1.2 Rumusan Masalah

Berikut merupakan rumusan masalah yang telah disusun pada penelitian:

1. Bagaimana kinerja algoritma *Random Forest*, *Decision Tree*, dan *KNN* pada sentimen analisis aplikasi IDN?
2. Bagaimana membuat model sentimen analisis dengan menggunakan metode CRISP-DM?

1.3 Batasan Masalah

Berikut merupakan batasan masalah dalam penelitian:

1. Penelitian ini menggunakan hasil data scraping dari ulasan aplikasi IDN pada *Google Playstore*
2. Data yang digunakan untuk penelitian sejumlah 995 data
3. Ulasan yang diambil untuk dataset hanya yang berbahasa Indonesia
4. Metode yang digunakan adalah *Random Forest*, *Decision Tree*, *KNN*

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Tujuan dari penelitian ini yaitu:

1. Sentimen analisis Terhadap ulasan Aplikasi IDN pada *Play store* dengan menggunakan metode *Random Forest*, *Decision Tree*, *KNN*.
2. Mengevaluasi tingkat akurasi Metode dalam melakukan sentimen analisis terhadap review aplikasi IDN pada *Play store*.

1.4.2 Manfaat Penelitian

Penelitian ini mempunyai manfaat, baik segi teoritis maupun praktis. Manfaat teoritis merupakan manfaat jangka panjang dalam pengembangan teori pembelajaran, sedangkan manfaat praktis memberikan dampak secara langsung terhadap komponen-komponen pembelajaran. Berikut manfaat teoritis pada poin 1, 2 dan manfaat praktis pada poin 3, 4 :

1. Mengetahui bagaimana merancang Sentimen analisis terhadap *Review Aplikasi IDN* dengan metode *Random Forest, Decision Tree, KNN*.
2. Mengetahui tingkat keakuratan algoritma *Random Forest, Decision Tree, KNN* dalam melakukan sentimen analisis terhadap review aplikasi IDN pada *Play store*.
3. Penelitian ini dapat digunakan calon pengguna untuk melakukan penentuan keputusan untuk *download* aplikasi IDN.
4. Penelitian ini dapat digunakan untuk developer aplikasi IDN untuk menentukan arah pengembangan aplikasi.

1.5 Sistematika Penulisan

Penelitian ini memiliki 5 bab yang terdiri dari:

1. BAB I Pendahuluan

Pada bab I menjelaskan mengenai latar belakang pada penelitian yang berjudul “Analisis Sentimen Terhadap Ulasan IDN News Menggunakan Komparasi *Random Forest, Decision Tree* dan *KNN*”, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

2. BAB II Landasan Teori

Pada bab II didalamnya terdapat pembahasan mengenai dasar teori yang digunakan dalam penelitian. Beberapa dasar teori yang akan di bahas pada bab ini yaitu Sentimen analisis, *Google Playstore, Random Forest, Decision Tree*, dan *KNN, Jupyter Notebook, Scraping, Data Preprocessing, TF-IDF, Phyton*.

3. BAB III Metodologi Penelitian

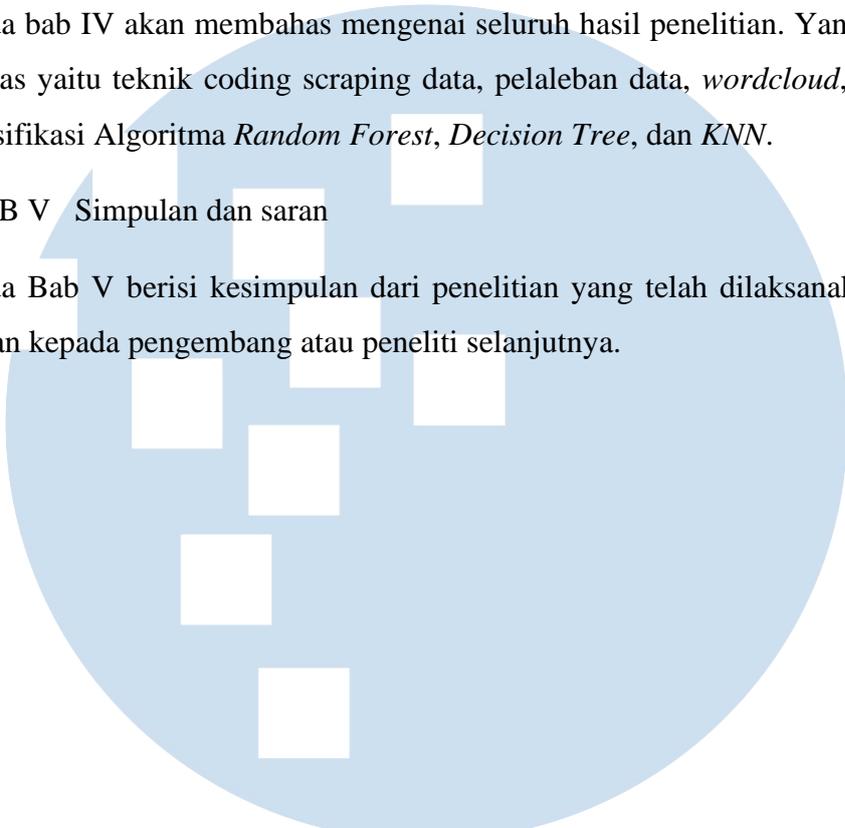
Pada bab III akan membahas mengenai Metodologi penelitian dan rancangan penelitian yang terdiri dari cara pengambilan data, alur penelitian.

4. BAB IV Hasil dan Pembahasan

Pada bab IV akan membahas mengenai seluruh hasil penelitian. Yang akan di bahas yaitu teknik coding scraping data, pelaleban data, *wordcloud*, *TF-IDF*, klasifikasi Algoritma *Random Forest*, *Decision Tree*, dan *KNN*.

5. BAB V Simpulan dan saran

Pada Bab V berisi kesimpulan dari penelitian yang telah dilaksanakan, serta saran kepada pengembang atau peneliti selanjutnya.



UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA