

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

*Fintech / financial technology* (teknologi keuangan) merupakan sebuah layanan keuangan yang menghasilkan model-model bisnis, aplikasi, maupun produk yang berkaitan dengan layanan keuangan dan perbankan. Salah satu produk yang dihasilkan dengan perkembangan teknologi keuangan ini adalah aplikasi dompet digital.

Dompet digital merupakan salah satu sistem aplikasi yang memanfaatkan teknologi internet yang bersifat online untuk melakukan berbagai macam transaksi yang berkaitan mengenai keuangan yang bersifat online. Dengan adanya aplikasi dompet digital di Indonesia akan berpotensi mendorong perkembangan pada *industry fintech* dikarenakan banyaknya jumlah masyarakat yang sudah tersambung pada internet dengan media *smartphone* dan juga letak geografis Indonesia sebagai negara kepulauan sehingga aplikasi dompet digital ini dapat mempermudah pengguna (*user*) dalam hal ini masyarakat dalam melakukan transaksi *online* yang berkaitan berkaitan mengenai keuangan.

Perkembangan pada *fintech / financial technology* (teknologi keuangan) juga sejalan dengan perkembangan internet di Indonesia berdasarkan riset dan survei yang dilakukan oleh Asosiasi Penyelenggara Jasa Internet (APJII) periode 2019 - 2020 tercatat bahwa kenaikan jumlah pengguna internet di Indonesia sangat signifikan sekali yaitu sebanyak 196,7 juta jiwa yang sudah tersambung internet (Katadata, 2020) dari total populasi manusia yang tercatat dalam survei kependudukan sebanyak 269,6 juta jiwa periode tahun 2020 (Kusnandar, 2020). Dengan adanya survei tersebut memberikan kesempatan kepada perusahaan-perusahaan aplikasi yang bergerak pada industri *financial technology* di Indonesia untuk terus melakukan inovasi teknologi guna meningkatkan pelayanan, sehingga pengguna dapat dengan mudah cepat dan aman dalam menggunakan aplikasi dompet digital. Dengan perkembangan, aplikasi dompet digital kini dapat diakses dengan menggunakan perangkat *android* maupun *ios* sehingga dapat dengan mudah diakses

oleh pengguna. Dengan segala potensi dan perkembangannya cukup tinggi khususnya di Indonesia membuat berbagai macam perusahaan tertarik dalam membuat dan mengembangkan aplikasi dompet digital salah satu perusahaan aplikasi dompet digital yang saat ini sedang berkembang di Indonesia yaitu aplikasi Dana (dompet digital Indonesia) yang dinaungi oleh PT. Espay Debit Indonesia Koe. Aplikasi tersebut sudah memiliki pengunduh dalam hal ini pengguna (*user*) pada situs *google play* sebanyak 10 juta lebih. Aplikasi Dana adalah sebuah layanan dompet digital Indonesia yang didesain untuk pembayaran transaksi secara non-tunai dan non kartu secara digital (*online*) dengan bantuan perangkat *smartphone* (DANA, 2021).

Aplikasi DANA juga memiliki sistem yang berbasis *open-platform* yang dapat tersambung dengan aplikasi sistem layananan digital lainnya yang sudah bermitra dengan DANA seperti berbagai macam restoran , BPJS, minimarket, supermarket dan 1000 merchant lainnya yang tersebar di seluruh Indonesia . Aplikasi ini diperkenalkan di publik pada 21 maret 2018 sebagai alat pembayaran (*mobile payment*) dan transfer antar bank yang aman, mudah dan cepat. Aplikasi Dana juga sudah mendapatkan perizinan dari Bank Indonesia (BI) dan terdaftar di DUKCAKPIL (Ditjen Kependudukan dan Pencatatan Sipil Kemendagri) sebagai lembaga teknologi financial Indonesia (Nirawati et al., 2020). Aplikasi DANA juga menjamin keamanan data penggunanya dengan menerapkan sistem authentication yang canggih dan otomatis seperti menggunakan kode pin OTP, dan *kode QR* yang bersifat dinamis.

*Google play* adalah sebuah situs layanan digital konten yang dibuat oleh google yang berisi berbagai macam produk-produk atau aplikasi digital yang berisi berbagai kategori seperti musik, buku, *mobile apps*, *games* dan masih banyak yang lainnya. Berbagai macam layanan tersebut dapat diakses melalui situs website atau aplikasi yang tersedia pada android. *Google play* mempunyai sebuah fitur yang berisikan ulasan penilaian (*review*) dari para penggunanya. Ulasan dari pengguna aplikasi tersebut biasanya menjadi tolak ukur yang efektif dan efisien untuk menemukan informasi terhadap produk, yang dapat digunakan untuk melihat ulasan dari pengguna aplikasi Dana. Ulasan atau *review* yang diberikan oleh pengguna aplikasi Dana pada umumnya berisikan keluhan yang bersifat positif maupun

negatif yang dituliskan secara tidak langsung baik banyak maupun sedikit, yang akan berpengaruh terhadap calon pengguna baru. Dari sekumpulan ulasan atau *review* yang cukup banyak, dari penilaian rating yang diberikan oleh pengguna akan berpengaruh terhadap calon pengguna baru (Santoso & Nurmalina, 2017). Oleh sebab itu, maka diperlukanya sebuah metode yang dapat memproses dan mengklasifikasikan data tersebut, yaitu teknik *scraping*. *scraping* merupakan teknik untuk mendapatkan informasi yang tersedia *website* secara otomatis tanpa harus menyalinnya secara manual.

Dalam penelitian ini, pengumpulan data yang berupa ulasan atau *review* yang diberikan pengguna kepada aplikasi Dana (dompet digital Indonesia) dengan menggunakan teknik *scraping*. Kemudian data akan memasuki proses analisis, penulis melakukan klasifikasi teks ulasan atau *review* yang guna untuk mengidentifikasi apakah ulasan tersebut bersifat positif atau negatif. Setelah melakukan proses klasifikasi, penulis akan mencoba mengekstrak dan mengeksplorasi yang seluas-luasnya mengenai informasi apa saja yang terdapat pada ulasan-ulasan yang ada serta sekiranya dianggap penting untuk digunakan pada berbagai keperluan, sehingga perusahaan dapat meningkatkan kualitas mutu produk dan kualitas pelayanan serta dapat meningkatkan pengguna dari aplikasi Dana. Dalam proses klasifikasi *sentiment* yang akan dilakukan dengan menggunakan metode *Naïve Bayes*.

Metode *Naïve Bayes Classifier* merupakan salah satu metode yang dapat digunakan untuk klasifikasi dan untuk melakukan pengelompokkan suatu opini dengan baik. *Naïve Bayes Classifier* mampu mengelompokkan komentar orang-orang menjadi positif ataupun negatif (M. W. A. Putra et al., 2020). Kemudian untuk melakukan proses ekstraksi dan eksplorasi penulis menggunakan statistik deskriptif dan asosiasi antar terms (topik yang sering dibicarakan dalam ulasan) yang saling berkaitan satu sama lain.

Oleh sebab itu dikarenakan tingkat akurasi yang cukup tinggi pada metode *Naïve Bayes Classifier* di antara metode lainnya, maka banyak peneliti yang menggunakan metode tersebut pada penelitiannya dalam hal ini yaitu menganalisis sentiment ulasan pada beberapa aplikasi dana, ovo, gopay, seperti pada penelitian sebelumnya yang sudah dilakukan diantaranya, yang dijelaskan pada **Tabel 1.1** sebagai berikut:

**Tabel 1. 1** Perbandingan hasil akurasi menggunakan beberapa metode

DANA	GOPAY	OVO
<p>1. Peneliti : Yogi Harfian (Harfian, 2021) Tahun 2021</p> <p>2. Metode : <b><i>Naïve Bayes</i></b> Menghasilkan tingkat akurasi sebesar <b>93,33%</b> dengan perbandingan <b>90% : 10%</b></p>	<p>1. Peneliti : Rachmat Mahendrajaya, Ghulam Asrofi Buntoro, Moh. Bhanu Setyawan (Mahendrajaya et al., 2019) Tahun 2019</p> <p>2. Metode : <b>Support Vector Machine (SVM)</b> Menghasilkan tingkat akurasi tertinggi sebesar <b>89,17%</b> Dari total dataset sebesar <b>1210 (923 positif 287 negatif)</b></p>	<p>1. Peneliti : Fajar Romadoni, Yuyun Umaidah, Betha Nurina Sari (Romadoni et al., 2020) Tahun 2020</p> <p>2. Metode : <b>Support Vector Machine (SVM)</b> Setelah hasil uji coba dengan beberapa perbandingan data uji dan data latih dengan perbandingan diantaranya adalah <b>60:40, 70:30, 80:20, 90:10</b> Akhirnya Menghasilkan tingkat akurasi Tertinggi <b>98,7%</b> Dari perbandingan sebesar <b>90:10</b> Dari total data sebesar <b>3852 (2034 pos, 1818 neg)</b></p>

1. Peneliti : Surohman, Sopian Aji, Rousyati, Fanny Fatma Wati (Surohman et al., 2020) Tahun 2020 2. Metode : <b>a. Naïve Bayes</b> menghasilkan tingkat akurasi sebesar <b>84,76 %</b> <b>b. K-NN</b> menghasilkan tingkat akurasi sebesar <b>82,92 %</b>		
---	--	--

Pada penjelasan **Tabel 1.1** diatas, terbukti bahwa metode algoritma *naïve bayes classifier* pada saat dibandingkan dengan metode algoritma yang lainnya, metode *naïve bayes classifier* memiliki tingkat akurasi yang cukup unggul dibandingkan dengan metode algoritma yang lainnya. Oleh sebab itu, peneliti memutuskan untuk menggunakan metode algoritma *Naïve Bayes Classifier (NBC)* dalam proses penelitian yang berjudul **“Klasifikasi Ulasan Pada Google Play Menggunakan Metode Naive Bayes Classifier (NBC) (Studi kasus aplikasi dompet digital DANA)”**.

## 1.2 Identifikasi Masalah

Berdasarkan latar belakang yang sudah dijelaskan sebelumnya dapat diidentifikasi masalahnya yaitu cukup banyaknya ulasan yang masuk pada kolom review aplikasi DANA tetapi, belum adanya permodelan klasifikasi yang menjelaskan mengenai ulasan review yang bersifat positif, negatif pengguna terhadap aplikasi DANA

## 1.3 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, terdapat permasalahan yang akan diangkat dalam skripsi ini yaitu sebagai berikut :

1. Bagaimana membuat pengklasifikasi yang dapat menganalisis tiap ulasan yang bersifat positif, negatif, yang datanya didapatkan langsung dari situs *Google play*
2. Bagaimana hasil dari penerapan metode *Naïve Bayes Classifier* dalam mengklasifikasikan data ulasan pengguna aplikasi *Dana* menjadi kelas positif dan negatif ?
3. Informasi topik apa yang diperoleh dalam proses klasifikasi dan asosiasi teks yang telah dilakukan ?

## 1.4 Batasan Masalah

Agar dalam proses penelitian lebih spesifik dan berfokus, oleh sebab itu maka perlunya ada pembatasan suatu masalah dalam proses penelitian. Agar tujuan dari penelitian dapat tercapai dan tersampaikan dengan baik, berikutlah batasan masalah yang akan dijabarkan :

1. Pada penelitian ini data yang akan digunakan adalah data dari ulasan komentar pengunduh pada aplikasi Dana yang tersedia pada situs resmi *googleplay*.
2. Data yang akan diteliti pada penelitian ini adalah data ulasan yang tercatat pada kolom ulasan aplikasi Dana pada *google play* per-periode 20 April

2021–25 Mei 2021 dan data yang diambil hanya data yang berbahasa Indonesia,

3. Penulis menggunakan bantuan *software Microsoft Excel 2016, Rstudio versi 4.0.4* lalu penulis juga menggunakan bantuan aplikasi ekstension yang disediakan pada *browser google chrome* untuk melakukan proses pengumpulan data menggunakan metode *web Scraping* dengan aplikasi *Data Miner*, .

### **1.5 Tujuan Penelitian**

Tujuan yang ingin dicapai dalam proses penelitian ini adalah :

1. Untuk mengidentifikasi dan menemukan suatu permasalahan yang berkaitan dengan bidang informatika sebagai bahan pembelajaran untuk menyelesaikan tugas akhir .
2. Mengimplementasikan Teknik *Web Scrapping* untuk mendapatkan data ulasan aplikasi dompet digital Dana dari situs Google play.
3. Memperoleh informasi yang penting dan berguna dengan cara menggunakan metode *naïve bayes classifer* dan asosiasi teks.
4. Mengimplementasikan metode *Naïve bayes classifier* dalam mengklasifikasi data ulasan pengguna aplikasi dompet digital Dana dari situs *Google play*.
5. Mengetahui tingkat ke akuratan dari metode algoritma *Naïve Bayes Classifier (NBC)* dalam memproses ulasan komentar yang ada pada aplikasi Dana di *google play* .
6. Untuk mengetahui faktor apa saja yang harus dilakukan untuk memperbaiki hasil ulasan negatif yang didapat.

### **1.6 Manfaat Penelitian**

Adapun manfaat yang ingin dicapai dan didapatkan dari proses penelitian kali ini adalah meliputi sebagai berikut :

1. Dapat mengetahui tingkat akurasi algoritma *Naïve bayes* dalam memproses data ulasan pengguna yang ada pada aplikasi Dana dalam bentuk teks.

2. Mengetahui gambaran secara umum mengenai ulasan pada aplikasi DANA berdasarkan data ulasan pada situs *Google play*.
3. Permodelan klasifikasi ulasan tentang aplikasi Dana ini Untuk memudahkan kepada pihak pengembang aplikasi maupun pemilik perusahaan aplikasi tersebut dalam mengetahui kepuasan pengguna dalam bentuk opini ulasan yang bersifat positif, negatif, sehingga dapat dijadikan acuan, masukan maupun referensi kepada perusahaan dalam upaya menjaga kualitas, meningkatkan kinerja dan memperbaiki kekurangan serta evaluasi agar aplikasi tersebut agar menjadi lebih baik kedepanya.

### **1.7 Tempat dan waktu penelitian**

#### **a. Tempat penelitian**

Tempat penelitian ini dilakukan pada situs resmi *google play* yang beralamat pada *website* berikut : [www.play.google.com](http://www.play.google.com), objek data yang akan diolah adalah data ulasan pengguna aplikasi dana pada situs *googleplay*.

Dana (Dompet Digital Indonesia) yang dinaungi oleh PT.Espay Debit Indonesia Koe yang beralamat di Capital Place lantai 18 Jl. Gatot Subroto, RT.6/RW.1. kelurahan Kuningan barat, kecamatan Mampang prapatan kota Jakarta selatan.

#### **b. Waktu penelitian**

Penelitian ini dilakukan dan dilaksanakan pada periode semester genap tahun ajaran 2020/2021 yang dimulai pada saat bulan maret 2021 sampai pada bulan juli 2021.

### **1.8 Sistematika Penulisan**

Sistematika pembahasan pada penyusunan skripsi ini dibagi menjadi 5 (lima) bab. Berikut penjelasan tentang masing masing bab.



## **BAB I PENDAHULUAN**

Pada bab ini berisi mengenai latar belakang penulisan skripsi, identifikasi masalah, rumusan masalah, Batasan masalah, tujuan penelitian, tujuan penelitian, manfaat penelitian, tempat dan waktu penelitian , metode penelitian dan sistematika penulisan.

## **BAB II LANDASAN TEORI**

Setidaknya berisi tentang Penelitian Sebelumnya berkaitan dengan topik skripsi yang dipilih. Jika dilengkapi dengan teori – teori pendukung lainnya lebih baik.

## **BAB III METODOLOGI PENELITIAN**

Setidaknya berisi tentang Kerangka Penelitian. Jika ditambahkan objek penelitian dan lainnya, lebih baik.

## **BAB IV PERANCANGAN SISTEM DAN IMPLEMENTASI**

Pada bab ini berisi berbagai macam penjelasan hasil dari penelitian dan hasil dari analisis penelitian yang dilakukan.

## **BAB V PENUTUP**

Pada bab ini berisi berbagai macam kesimpulan hasil dari penelitian yang dilakukan berupa keterbatasan sistem serta saran yang diberikan oleh peneliti kepada pembaca jika ingin melanjutkan penelitian dengan tema terkait.

## **BAB II**

### **LANDASAN TEORI**

#### **2.1 Penelitian terkait**

Penelitian yang menjelaskan mengenai pengolahan data dan pengklasifikasian sekumpulan ulasan sentimen yang bertujuan guna mendapatkan sebuah informasi, sudah cukup banyak dilakukan oleh peneliti-peneliti sebelumnya. Penelitian terdahulu tersebut cukup penting bagi saya selaku penulis sebagai sarana kajian agar mengetahui keterkaitan antara penelitian terdahulu dengan penelitian yang akan saya lakukan. Tujuan dari adanya tinjauan pustaka ini adalah sebagai sarana untuk menunjukkan bahwa penelitian terdahulu sangatlah bermanfaat dalam perkembangan ilmu pengetahuan dan juga untuk sebagai pembandingan dengan penelitian yang akan saya lakukan. Berikut adalah beberapa tinjauan penelitian terdahulu yang pernah dilakukan terkait penelitian yang akan saya lakukan, beberapa jurnal dan penelitian tersebut dijadikan oleh penulis sebagai acuan.

Penelitian mengenai komentar aplikasi Dana pada Instagram sebelumnya pernah dilakukan oleh (Harfian, 2021), dalam tugas akhirnya yang berjudul “Klasifikasi sentiment aplikasi dompet digital Dana pada komentar di Instagram menggunakan *Naïve Bayes Classifier*”. Penelitian tersebut menjelaskan mengenai pengolahan ulasan pengguna pada kolom komentar Instagram *official* Dana yang bertujuan untuk sebagai masukan bagi pemilik perusahaan mengenai keluhan dan permasalahan pengguna pada saat menggunakan aplikasi Dana, penelitian tersebut dilakukan menggunakan metode algoritma *Naïve bayes classifier* dan menggunakan Bahasa pemrograman PHP, lalu diuji coba menggunakan metode *whitebox* dan *confusion matrix* sehingga mendapatkan tingkat akurasi pada saat memproses data sebesar 93,33% dengan menggunakan perbandingan data latih sebesar 90% dan data uji sebesar 10% . hasil penelitian tersebut dapat dijadikan acuan untuk perusahaan untuk meningkatkan dan menjaga kualitas pada aplikasi .

Shinta prima astuti (Astuti, 2020) pada tugas akhirnya melakukan penelitian mengenai ulasan yang dilakukan pengguna aplikasi Tokopedia yang didapatkan pada situs *google play* dengan cara menggunakan teknik *Scraping*, ulasan komentar

tersebut berisi berbagai macam hal topik pembicaraan seperti keluhan pengguna pada saat menggunakan aplikasi tersebut ataupun ulasan mengenai kepuasan pengguna pada saat berbelanja, dll. Penelitian tersebut menggunakan metode *Naïve bayes classifier* dan dikelompokkan menjadi beberapa klaster yaitu, kebermanfaatan, pelayanan, pengalaman belanja dan tampilan aplikasi. Beberapa hasil klaster tersebut dilakukan proses *resampling* dengan teknik *Random Over Sampler* dan *Smoteenn* pada kurva ROC dan AUC, lalu setelah itu diproses dan menghasilkan akurasi tertinggi sebesar 92,5% dan nilai AUC sebesar 0,95.

Penelitian yang dilakukan (Surohman et al., 2020) yang menghasilkan sebuah jurnal yang berjudul “Analisa Sentimen Terhadap Review *Fintech* Dengan Metode *Naive Bayes Classifier* Dan *K- Nearest Neighbor*” pada penelitian ini meneliti mengenai aplikasi *fintech* yang sedang marak digunakan yaitu Dana dikarenakan aplikasi tersebut menyediakan kemudahan, keamanan, dan banyaknya promosi. Dikarenakan maraknya aplikasi Dana digunakan oleh masyarakat, sehingga melandasi dilakukannya penelitian tersebut yang bertujuan sebagai saran, masukan kepada pemilik dan pengembang aplikasi untuk meningkatkan kualitas produk aplikasi. Pada penelitian tersebut data yang akan diproses didapatkan dari situs *Google play store* dan diklasifikasikan menjadi positif dan negatif lalu diproses menggunakan kedua metode yang berbeda, pada saat menggunakan metode *naïve bayes classifier* mendapatkan tingkat akurasi sebesar 84,72% sedangkan pada saat menggunakan metode K-NN mendapatkan tingkat akurasi lebih rendah yaitu sebesar 82,92%.

Penelitian yang dilakukan Zia Ayu Nuansa Gumilang (Gumilang, 2018) pada tugas akhirnya adalah melakukan penelitian mengenai pengolahan sentimen ulasan pengguna pada saat menggunakan aplikasi shopee, peneliti melakukan penelitian tersebut dilandaskan karena aplikasi *e-commerce* tersebut memiliki reputasi yang cukup baik di Indonesia. Pada penelitian tersebut peneliti mendapatkan data ulasan secara langsung pada situs *google play*. Setelah itu data tersebut dianalisis dan diberi pelabelan menggunakan metode *naïve bayes classifier* untuk diklasifikasikan ulasan tersebut berdasarkan kategori positif dan negatif menghasilkan tingkat akurasi sebesar 97,4%. Setelah diklasifikasikan perkategori data tersebut selanjutnya, ulasan tersebut diasosiasikan perkelas guna untuk mendapatkan

informasi yang dianggap penting, berguna dalam proses pengambilan keputusan dan dalam pemecahan masalah.

Penelitian yang dilakukan (Romadoni et al., 2020) pada jurnalnya yang berjudul “*Text Mining Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan Algoritma Support Vector Machine*” jurnal tersebut menjelaskan mengenai aplikasi dompet digital ovo yang cukup banyak digemari di Indonesia. Jurnal penelitian tersebut meneliti tentang opini pengguna aplikasi dompet digital ovo pada sosial media twitter yang bertujuan untuk mengklasifikasikan ulasan pengguna tersebut secara positif maupun negatif. Penelitian tersebut mendapatkan data secara langsung dengan Teknik web scarping pada sosial media twitter dengan kata kunci @ovo\_official dan mendapatkan total data sebanyak 3852 tweet dengan rincian ulasan positif sebanyak 2034, sedangkan ulasan negatif sebanyak 1818. Hasil klasifikasi ulasan tersebut lalu data tersebut dilakukan skenario *splitting data* menggunakan algoritma *Support vector machine (SVM)* dengan beberapa rasio yaitu, 60:40, 70:30, 80:20, 90:10 dan dengan 4 kernel yaitu kernel linear, kernel rbf, sigmoid dan polynomial. Setelah dilakukan uji coba beberapa skenario yang menghasilkan perbandingan tingkat akurasi tertinggi adalah 90:10 yang mendapatkan nilai akurasi 98.7%.

Penelitian selanjutnya dilakukan oleh Rachmat Mahendrajaya, Ghulam Asrofi Buntoro, Moh. Bhanu Setyawan (Mahendrajaya et al., 2019) dalam jurnalnya yang membahas dan memproses data ulasan komentar keluhan pengguna aplikasi *fintech* Gopay. Pada penelitian tersebut dataset didapatkan sebanyak 1210 ulasan secara langsung pada sosial media twitter dari data tersebut. Lalu setelah itu, masuk pada tahapan preprocessing dan dilabeli menjadi ulasan positif sebanyak 923, ulasan negatif sebanyak 287 menggunakan metode *lexicon based*. Hasil pelabelan tersebut diklasifikasikan dan diproses menggunakan metode *support vector machine (SVM)* dengan kernel linear yang menghasilkan tingkat akurasi sebesar 89,17% sedangkan pada saat menggunakan kernel polynomial menghasilkan tingkat akurasi sebesar 84,38%.

Penelitian yang akan dilakukan adalah mengimplementasikan teknik *web scraping* untuk mengumpulkan data ulasan pengguna aplikasi dompet digital Dana dari situs *Google play* untuk kemudian dilakukan analisis klasifikasi sentimen positif dan negatif menggunakan metode *Naïve Bayes Classifier*. Setelah diperoleh hasil klasifikasi, kemudian dilakukan proses visualiasasi, eksplorasi dan ekstraksi informasi dengan pendekatan *text mining*. Pada Tabel 2.1 menjadi perbandingan penelitian sebelumnya dan penelitian yang akan dilakukan oleh penulis.

**Tabel 2. 1 Penelitian Terdahulu Terkait Penelitan**

No	Nama Penulis	Judul	Tahun	Metode	Hasil
1	Yogi Harfian (Harfian, 2021)	Klasifikasi sentimen aplikasi dompet digital Dana pada komentar di Instagram menggunakan <i>Naïve Bayes Classifier</i>	2021	<i>Naïve Bayes Classifier</i> (NBC)	Hasil dari klasifikasi penelitian kali ini memiliki tingkat akurasi tertinggi mencapai 93,33% dan pada pengujian menggunakan data latih dan data uji sebesar 90% : 10%
2	Shinta Prima Astuti (Astuti, 2020)	Analisis sentimen berbasis aspek pada aplikasi TOKOPEDIA menggunakan <i>LDA &amp; Naïve Bayes</i>	2020	<i>LDA &amp; Naïve Bayes</i>	Penelitian ini menghasilkan tingkat akurasi sebesar 92,5%. Sedangkan untuk hasil nilai dari AUC menunjukan nilai yang cukup baik yaitu sebesar 0,95.

3	Surohman, Sopian Aji, Rousyati, Fanny Fatma Wati (Surohman et al., 2020)	Analisa Sentimen Terhadap <i>Review Fintech</i> Dengan Metode <i>Naive Bayes Classifier</i> Dan <i>K-Nearest Neighbor</i>	2020	<i>Naive Bayes Classifier &amp; K-Nearest Neighbor</i>	Tingkat akurasi yang didapatkan dari penelitian kali ini adalah sebesar 84,76% +/-3,93% menggunakan metode <i>Naive Bayes</i> sedangkan yang menggunakan metode <i>KNN</i> menghasilkan tingkat akurasi lebih rendah yaitu sebesar 82,92% +/-4,87%
4	Zia Ayu Nuansa Gumilang (Gumilang, 2018)	Implementasi <i>Naive Bayes Classifier</i> dan asosiasi untuk Analisa sentiment data ulasan aplikasi <i>E-Commerce SHOPPE</i> pada situs <i>Google play</i>	2018	<i>Naive Bayes Classifier (NBC)</i>	Hasil dari penelitian ini menghasilkan tingkat akurasi sebesar 97,4% dengan perbandingan data latih dan data uji sebesar 80%   20% <sub>s</sub>
5	Fajar Romadoni, Yuyun Umaidah, Betha Nurina Sari	<i>Text Mining</i> Untuk Analisis Sentimen Pelanggan Terhadap Layanan Uang Elektronik Menggunakan	2020	<i>Support Vector Machine (SVM)</i>	Hasil dari penelitian ini menghasilkan tingkat akurasi sebesar 98,7%. Dengan total <i>dataset</i> yang

	(Romadoni et al., 2020)	Algoritma <i>Support Vector Machine</i>			digunakan sebanyak 3852 <i>tweet</i> yang diambil dengan Teknik <i>web scraping</i> dari twitter dengan kata kunci @ovo_id , dengan beberapa perbandingan data uji dan data latih sebesar 60:40, 70:30, 80:20, 90:10 dan hasil perbandingan tertinggi pada 90:10
6	Rachmat Mahendrajaya, Ghulam Asrofi Buntoro, Moh. Bhanu Setyawan (Mahendrajaya et al., 2019)	Analisis Sentimen Pengguna GOPAY Menggunakan Metode <i>Lexicon Based</i> dan <i>Support Vector Machine</i>	2019	Metode <i>Lexicon Based &amp; Support Vector Machine (SVM)</i>	penelitian ini menghasilkan tingkat akurasi sebesar 89,17% dengan kernel linear dan 84,38% dengan menggunakan kernel <i>Polynomial</i> Dengan total dataset yang digunakan sebanyak 1210 (923 <i>tweet</i> positif, 287 negatif <i>tweet</i> )

7	Siti Masturoh, Achmad Baroqah Pohan (Masturoh & Pohan, 2020)	<p><i>SENTIMENT ANALYSIS AGAINST THE DANA E-WALLET ON GOOGLE PLAY REVIEWS USING THE K-NEAREST NEIGHBOR ALGORITHM.</i></p> <p>ANALISIS SENTIMEN TERHADAP APLIKASI DOMPET DIGITAL DANA PADA <i>GOOGLE PLAY</i> MENGGUNAKAN ALGORITMA <i>K-NEAREST NEIGHBOR</i> (KNN)</p>	2020	<i>K_Nearest Neighbor (KNN)</i>	<p>Penelitian ini menghasilkan tingkat akurasi sebesar 86,64% dari nilai pengujian 3 kelas berdasarkan pemberian bintang, <b>kelas pertama</b> terdiri dari bintang 1-5, <b>kelas ke 2</b> terdiri dari (bintang 1 &amp; 5, <b>kelas ke 3</b> terdiri dari pemberian label pada bintang (1 &amp; 2 label negatif, 3 label netral, serta bintang 4 &amp; 5 label positif) dan pengujian nilai k 1-10, sehingga didapatkan nilai accuracy tertinggi dengan kelas ke 2 (bintang 1 dan bintang 5)</p>
---	--	--	------	---------------------------------	---



## 2.2 Google play

*Google play* (yang dulunya bernama android market) adalah sebuah situs layanan digital konten yang dibuat oleh google yang berisi berbagai macam produk-produk atau aplikasi digital yang berisi berbagai kategori seperti musik, buku, *mobile apps*, *games* dan masih banyak yang lainnya (update per tanggal 10 Desember 2017), sejumlah 3.532.448 aplikasi. Banyaknya aplikasi yang terdapat pada *google play* membuat toko aplikasi ini sangat menarik untuk dijadikan sebagai objek penelitian, khususnya dalam bidang data mining. Salah satu contohnya yaitu analisis sentimen yang merupakan cabang ilmu dari data mining, sangat cocok jika diaplikasikan pada *google play*. Analisis sentimen digunakan untuk menentukan sentimen para pengguna dari setiap aplikasi yang terdapat pada toko aplikasi tersebut dengan mengklasifikasikan ratusan, bahkan ribuan *text review* dari pengguna secara otomatis. (Ilmawan, 2018)

*Google play* juga selain dapat diakses melalui situs berbasis website, bisa juga diakses melalui aplikasi android/ios . *Google play* juga dilengkapi dengan fitur ulasan atau penilaian (*review*) pada sebuah aplikasi atau produk digital yang berfungsi untuk menemukan informasi aplikasi tersebut dalam perspektif pengguna yang sudah menggunakan aplikasi tersebut (Aaputra, 2019).

Dapat disimpulkan bahwa *google play* adalah salah satu *platform digital* yang berisi berbagai macam aplikasi digital konten yang dibuat oleh google dan dilengkapi berbagai macam fitur, salah satunya adalah fitur ulasan untuk menilai aplikasi apa situs *google play*. Oleh sebab itu dikarenakan banyaknya aplikasi dan ulasan pada aplikasi tersebut, menurut peneliti sangat cocok untuk dilakukan penelitian analisis sentiment mengenai ulasan sentimen pengguna aplikasi apa situs *google play*.

## 2.3 E-Wallet

*E-Wallet* (Dompet elektronik) adalah salah satu inovasi dari teknologi yang sangat pesat untuk melakukan transaksi jual beli ataupun perbankan . Dengan layanan dompet elektronik, pengguna dapat melakukan transaksi secara nontunai tanpa kartu baik *online* maupun *offline*. Di Indonesia sendiri banyak layanan

dompet digital atau *fintech* sebut saja GoPay, OVO, T-cash, Dana dan sebagainya (Mahendrajaya et al., 2019). Penggunaan dompet elektronik untuk penggunaan proses transaksi perniagaan di Indonesia menunjukkan perkembangan yang meningkat karena dinilai lebih praktis dan memiliki mobilitas yang tinggi. Banyaknya keuntungan yang ditawarkan oleh dompet elektronik yang berdasarkan fungsinya mirip seperti dompet pada umumnya untuk menyimpan uang ini seperti diskon dan cashback membuat masyarakat makin menggemari penggunaan dompet elektronik (M. W. A. Putra et al., 2020).

*E-Wallet* (Dompet Elektronik) juga menjamin sebuah keamanan informasi pengguna pada saat menggunakan *E-Wallet* dengan mengenkripsi informasi pengguna, dengan memberikan kode otentifikasi serta memberikan PIN (kata sandi) ketika ingin menggunakan perangkat tersebut. *E-Wallet* juga dapat dikoneksikan dengan semua bank guna untuk mempermudah dalam melakukan semua transaksi layanan yang bersifat *online* hanya dengan sekali klik pada perangkat *e-wallet* tersebut. (Upadhayaya, 2012)

Uang elektronik adalah alat pembayaran nontunai atau *cashless* yang nilai uangnya disimpan secara elektronik di dalam media server ataupun chip yang dapat dipindahkan untuk kepentingan transaksi pembayaran atau transfer dana (Romadoni et al., 2020)

## **2.4 Dana**

Aplikasi Dana (layanan dompet digital Indonesia) yang dinaungi oleh PT. Espay Debit Indonesia Koe yang berlokasi di Capital Place lantai 18 Jl. Gatot Subroto, RT.6/RW.1. kelurahan Kuningan barat, kecamatan Mampang prapatan kota Jakarta selatan ini adalah sebuah aplikasi berbasis digital (*online*) yang didesain untuk dijadikan sebagai pembayaran (*mobile payment*) non-tunai dan non-kartu yang dapat berjalan dengan cepat, praktis, mudah dan aman. Aplikasi Dana juga sudah tersedia pada *platform* android dan ios dapat unduh dan didapatkan secara gratis pada *Google play* atau *apps store*. Aplikasi Dana juga sudah bersifat *open-platform* yang dapat tersambung dengan beberapa aplikasi yang sudah bermitra dengan Dana, sehingga aplikasinya dapat digunakan untuk berbagai

macam transaksi yang bersifat *online*, seperti bayar tagihan listrik, *Top Up* saldo, Transfer antar bank, Setor tunai ke minimarket, Beli tiket bioskop, beli pulsa, beli makanan dan masih banyak yang lainnya (DANA, 2021). Aplikasi Dana juga telah memiliki izin legal dari Bank Indonesia sebagai Lembaga teknologi yang berja di dalam bidang *financial technology* di Indonesia. Selain itu aplikasi Dana juga memiliki beberapa fitur layanan keunggulan yaitu sebagai berikut (Nirawati et al., 2020) :

1. Proses lebih cepat dan mudah

Salah satu keunggulan fitur dalam proses pembayaran khususnya kepada pengguna ios adalah setiap tagihan, bukti pembayaran dan lainnya selalu tersimpan dan terupdate pada penyimpanan yaitu *icloud*, sehingga ketika pengguna memiliki saldo pada aplikasi Dana yang terisi cukup maka sistem akan menarik saldo tersebut dengan otomatis untuk membayarkan tagihan sesuai dengan detail tagihanya. Sehingga hal itu dapat mempermudah pengguna .

2. Memiliki berbagai macam promo

Salah satu fitur keunggulan yang ditawarkan oleh aplikasi Dana adalah berbagai macam promo menarik yang diberikan kepada pengguna seperti voucher makanan, voucher belanja dan lainnya dengan harga sangat murah .

3. Transfer lebih efisien

Seperti yang sudah dijelaskan sebelumnya bahwa Dana adalah aplikasi yang bersifat *open-platform* sehingga dapat terkoneksi dengan aplikasi atau perusahaan yang sudah bermitra dengan Dana . Aplikasi Dana juga sudah berkerjasama atau bermitra dengan berbagai macam bank yang ada di Indonesia, sehingga aplikasi Dana cukup diminati pengguna masyarakat dikarenakan dapat digunakan untuk tranfer antar bank gratis hanya dengan minimal 50.000., hal itu sehingga cukup mempermudah masyarakat (Nirawati et al., 2020).

## 2.5 Data tidak terstruktur

*Unstructured Data* (Data tidak terstruktur) adalah sebuah data yang tidak memiliki permodelan yang stabil atau tetap dan tidak beraturan. *Unstructured data* (data tidak terstruktur) jika tidak melalui tahapan *preprocessing* maka sekumpulan data tidak beraturan atau tidak terstruktur tersebut tidak akan bisa disimpan dan diproses kedalam suatu tabel, seperti contoh pada media sosial (*twitter, blog, facebook, dan masih banyak yang lainnya*), data *call center, email*, dan lain-lain. Data tidak terstruktur sangat berkaitan sekali dengan tiga *V Big Data* , yaitu diantaranya sebagai berikut :

1. *Volume* : Data tidak terstruktur (*Unstructured Data*) lebih memerlukan banyak ruang penyimpanan dibandingkan dengan data yang terstruktur
2. *Velocity* : Biasanya data yang tidak terstruktur (*Unstructured Data*) peningkatan datanya lebih cepat dibandingkan data yang terstruktur.
3. *Variety* : Biasanya data yang tidak terstruktur (*Unstructured Data*) sumber datanya didapatkan dari sumber data yang beum sama sekali masuk tahapan pemrosesan sehingga dapat mengungkapkan sebuah informasi pengguna yang cukup penting (Kurniasari, 2018).

## 2.6 JavaScript

*JavaScript* merupakan bahasa *script popular* yang digunakan untuk menciptakan Web yang dapat mampu berinteraksi dengan pengguna dan dapat merespon *event* yang terjadi pada suatu halaman. *JavaScript* merupakan sebuah perekat yang menyatukan halaman web. *Java Script* dikembangkan oleh Brendan Eich di Netscape pada tahun 1995 yang menjadi bagian Netscape 2.0. *JavaScript* tidak dirancang untuk menulis dan menyimpan *file* pada *server*. *JavaScript* digunakan untuk mendeteksi dan beraksi terhadap *event-event* yang disebabkan oleh pengguna . *JavaScript* berperan sebagai bahasa pemrograman yang memiliki konstruksi dasar variabel dan tipe data, *look control, statement if/else, statemen switch* fungsi serta objek, perhitungan aritmatika, memanipulasi tanggal dan waktu, string serta objek, memodifikasi array. (Sianipar, 2015)

## 2.7 Json (JavaScript Object Notation)

JSON yang merupakan singkatan dari *JavaScript Object Notation* adalah standar berbasis teks untuk pertukaran data. Format JSON dikenal ringan (berukuran kecil), mudah untuk dibaca, ditulis, dan dipahami manusia serta mudah untuk diuraikan dan dibuat oleh mesin. Format ini dibuat berdasarkan bahasa pemrograman *JavaScript*, standar ECMA-262 edisi ketiga – Desember 1999. JSON bersifat language independent namun menggunakan kaidah penulisan yang dikenal luas oleh programmer dari keluarga bahasa C (C, C++, C#, Java, JavaScript, Perl, Python, dll) (Wiyono & Wijayanto, 2012).

Penggunaan format *Javascript Object Notation (JSON)* bertujuan untuk memperkecil jumlah kolom yang akan digunakan pada basisdata agar tabel basisdata lebih ringkas dan efisien, selain itu format JSON merupakan format yang ringan dan mudah dibaca oleh berbagai bahasa pemrograman (Wisnu et al., 2019)

## 2.8 CSS (Cascading Style Sheet)

CSS merupakan salah satu *style sheet language* atau bahasa web yang berfungsi sebagai pengontrol format tampilan dari suatu halaman web yang ditulis dengan kode atau *mark up language*. Pada awalnya, CSS dipakai untuk merancang halaman HTML dan XHTML. Namun kini, pengaplikasian CSS sudah untuk semua dokumen XML, termasuk XUL, SVG, dan juga Android. CSS dipakai sebagai pemisah konten utama dengan tampilan dokumen.

CSS (*Cascading Style Sheet*) adalah sebuah cara untuk memisahkan isi dengan layout dalam halaman-halaman web yang dibuat. *Cascading Style Sheet* dikembangkan untuk menata tata pengaturan halaman web. Pada awalnya CSS dikembangkan pada SGML pada tahun 1970 dan terus dikembangkan hingga saat ini CSS telah mendukung banyak bahasa, *Cascading Style Sheet* memiliki arti gaya menata halaman bertingkat. Yang berarti setiap satu elemen yang telah format, maka anak dari elemen tersebut secara otomatis mengikuti format elemen induknya. (Fatmawati, 2016)

CSS digunakan untuk mengubah tampilan yang dihasilkan oleh tag HTML karena bekerja dengan kode yang berbeda (M. Y. Putra et al., 2020). Tampilan dokumen itu mencakup di dalamnya warna *font* serta *layout*. Pemisahan tersebut bisa membuat daya dalam spesifikasi tampilan meningkat. Sehingga kerumitan struktur konten dan kerumitan penulisan kode bisa dikurangi. Di samping itu juga membuka peluang bagi pembagian halaman dalam formatting. Sebagai contoh, penggunaan teknik *tableless* pada rancangan web. Tidak hanya itu, dengan adanya CSS sebuah halaman juga bisa ditampilkan dalam bermacam-macam model/ style dengan pembawaannya masing-masing. Misalnya *by voice*, *in-print*, *on-screen*, dan lain sebagainya. Pemilik web pun dapat menentukan link penghubung kontennya dengan file CSS. (Kezaldo, 2021)

## 2.9 JQuery

*Jquery* adalah salah satu teknik atau kumpulan *library javascript* yang sangat terkenal animasinya. *jquery* adalah *javascript library*, *jquery* mempunyai semboyan “*write, less, do more*”. *Jquery* dirancang untuk memperingkas kode-kode *javascript*. *Jquery* adalah *javascript* yang cepat dan ringan untuk menangani dokumen HTML, menangani event, membuat animasi dan *interaksi ajax*. *Jquery* dirancang untuk mengubah cara anda menulis *javascript* (Fatmawati, 2016)

## 2.10 Database My Sql

*Database* merupakan sebuah aplikasi yang berfungsi untuk penyimpanan data. Masing-masing database memiliki sistem perintah tertentu dalam pengoperasiannya. Baik itu dalam pembuatannya, pengaksesannya, maupun untuk pencarian serta penyalinan datanya. (Kezaldo, 2021)

*Database* secara sederhana adalah sebuah gudang penyimpanan data. Secara teori database adalah kumpulan data atau informasi yang kompleks lalu tiap data-data tersebut disusun menjadi beberapa kelompok dengan tipe data yang sejenis disebut *table/entity*), di mana setiap datanya dapat saling berhubungan satu sama lain atau dapat berdiri sendiri, sehingga mudah diakses. Sedangkan *My SQL*

merupakan sebuah *MySQL* merupakan database yang awalnya hanya berjalan pada sistem *Unix* dan *Linux*. Seiring berjalannya waktu dan banyaknya peminat yang menggunakan database ini, *MySQL* merilis versi yang dapat diinstal pada hampir semua *platform*, termasuk *Windows*.

*MySQL* merupakan sebuah manajemen berbasis data (*database*) yang biasa dipakai dalam pengelolaan data secara gratis. Dalam pengoperasiannya *MySQL* memakai bahasa perintah dasar yang cukup familiar dalam menghubungkan antara *software*-nya dengan *database server*, yaitu *SQL* atau *Structured Query Language*. DBMS atau *Database Management System* multi alur dan multi pengguna yang telah dipakai oleh *MySQL* ialah *database management system* yang open source dengan dua bentuk lisensi, yakni perangkat lunak bebas dan perangkat lunak berpemilik yang pemakaiannya terbatas. (Kezaldo, 2021)

Lisensi dari *MySQL* adalah freeware. Kita dapat mendownload dan menggunakannya tanpa harus membayarnya. Meskipun kita menjual produk menyertakan *software MySQL*, kita tidak melanggar hak cipta. Mungkin bagi kita yang baru dengan *MySQL* akan bingung dengan dua kata "*SQL*" dan "*MySQL*". Pertanyaan yang mungkin muncul adalah, apakah *SQL* itu sebenarnya, dan apa bedanya dengan *MySQL*?

*SQL* merupakan kependekan dari kata "*Structured Query Language*". *SQL* merupakan suatu bahasa permintaan yang terstruktur yang melekat pada satu database atau *SMBD* tertentu, sedangkan *MySQL* merupakan databasenya. Dengan kata lain, *MySQL* merupakan *SMBD*-nya dan *SQL* adalah perintah atau bahasa yang melekat di dalam *SMBD* tersebut. (Febio, 2011)

## **2.11 Data mining**

Data Mining merupakan ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari himpunan data yang banyak. Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik-teknik, metode-metode, atau algoritma dalam data mining sangat bervariasi. Pemilihan metode atau algoritma

yang tepat sangat bergantung pada tujuan dan proses *Knowledge Discovery in Database (KDD)* secara keseluruhan (Hayuning tyas, 2020).

*Knowledge Discovery in Databases* merupakan metode untuk mendapatkan pengetahuan dari basis data yang ada. Didalam basis data tersebut terdapat tabel-tabel yang saling berelasi. Hasil pengetahuan dalam proses tersebut dapat digunakan sebagai knowledge base untuk keperluan pengambilan keputusan. KDD dan data mining sering digunakan untuk penggalian informasi tersembunyi dari basis data yang besar (Romadoni et al., 2020). berikut adalah langkah-langkah dalam proses KDD :

1. *Data cleaning* Proses membersihkan data untuk menghilangkan noise dan data yang tidak konsisten.
2. *Data integration* Proses mengombinasikan data apabila memiliki sumber data yang masih terpecah dalam sistem data mining.
3. *Data selection* Pengambilan data yang sesuai untuk digunakan dalam proses analisis data mining.
4. *Data transformation* Proses dimana data ditransformasikan menjadi bentuk-bentuk yang sesuai untuk proses dalam data mining.
5. *Data mining* Proses untuk mengekstrak kumpulan/ himpunan data agar menjadi ilmu pengetahuan.
6. *Pattern evaluation* Proses untuk mengidentifikasi kebenaran dari pola data yang mewakili pengetahuan berdasarkan beberapa tindakan.
7. *Knowledge representation* Proses untuk menyajikan presentasi teknik visual digunakan untuk menampilkan hasil pengetahuan kepada pengguna.

Data mining juga memiliki banyak sekumpulan algoritma/metode/teknik pengumpulan dan pencarian data yang memiliki fungsi serta tujuan yang berbeda-beda. Oleh sebab itu data mining dapat dikelompokkan sebagai berikut :

1. Deskripsi

Tujuan dari proses deskripsi ini adalah guna untuk menemukan pola data yang sering muncul lalu dari data tersebut diubah menjadi pola yang dapat mempermudah suatu aktivitas. Seperti contoh algoritma yang tergolong dalam deskripsi adalah algoritma apriori.



## 2. Klasifikasi

Klasifikasi data mining adalah pengelompokan sebuah data berdasarkan variable kriteria dengan variable target. Seperti contoh algoritma yang tergolong dalam klasifikasi data mining adalah algoritma *Cart*, *ID3*, *C4.5*, *J48*, *C5.0*, *Naïve bayes*, *K-Nearest Neighbor* dan masih banyak yang lainnya.

## 3. Prediksi

Tujuan dari prediksi adalah untuk memprediksi sekumpulan nilai yang nantinya hasil dari prediksi tersebut dapat digunakan untuk membantu dalam pengambilan suatu keputusan dimasa yang akan datang. Seperti contoh algoritma yang tergolong dari prediksi adalah *ID3*, *J48*, *C5.0*, *Rough Set*, *Cart*.

## 4. Estimasi

Tujuan dari estimasi adalah untuk memperkirakan sebuah estimasi dengan cara klasifikasi dan pengelompokan yang dimana hasil estimasi tersebut lebih ke arah nilai yang numerik bukan ke arah yang bersifat kategori. Seperti contoh algoritma yang tergolong dalam estimasi adalah Regresi linear dan Regresi linear berganda dan masih banyak yang lainnya.

## 5. Pengklasteran

Pengklasteran merupakan sekumpulan klasifikasi data yang memiliki kesamaan dalam bentuk nilai (homogen) yang nantinya hasil dari pengklasteran tersebut memiliki bentuk data seperti, record data, kelas-kelas, objek-objek yang memiliki kesamaan. Dalam pengklasteran ini juga memiliki perbedaan dengan klasifikasi, hal yang membedakan itu adalah dari variable target/keputusanya. Seperti contoh algoritma yang tergolong dalam pengklasteran adalah *K-Medoids*, *K-Means*, *K-Nearest Neighbor* dan masih banyak yang lainnya.

## 6. Asosiasi

Asosiasi memiliki artian sebagai sebuah keeratan hubungan antara dua atau lebih variabel yang sifatnya kualitatif. Asosiasi merupakan sebuah proses pencarian hubungan antar elemen data. Sering disebut *market basket analysis* dalam dunia retail (Miner et al., 2012).

## 2.12 Analisis Deskriptif

Analisis deskriptif merupakan suatu teknik penggambaran suatu masalah berdasarkan kumpulan data dengan cara menyusun data dengan cara sedemikian rupa sehingga bisa dipahami dengan mudah karakteristik datanya yang nantinya akan berguna untuk proses selanjutnya, oleh sebab itu maka terdapat suatu proses pengumpulan dan pengolahan data berdasarkan tujuannya. Pengolahan data pada analisis deskriptif hanya berhubungan dalam menguraikan atau memberikan keterangan pada data yang secara umum (Fikria, 2018).

Analisis data dengan menerapkan metode deskriptif dinyatakan sebagai analisis statistik sederhana atau yang paling sederhana. Akan tetapi, hasil analisis statistik deskriptif tersebut dapat menjadi masukan yang sangat berharga untuk para pengambil keputusan, tergantung pada bentuk dan cara menyajikan hasil analisis tersebut (Agung, 2016) Tujuan dari analisis deskriptif adalah sebagai berikut :

- a. Untuk menilai atau mengevaluasi, apakah data yang dipakailayak dapat dipercaya atau tidak.
- b. Untuk mempelajari perbedaan nilai statistik variabel-tujuan seperti prevalensi, proporsi dan rata-rata disertai dengan standar deviasinya yang dihitung berdasarkan data sampel tertentu dengan nilai yang diharapkan. Dengan kata lain, menentukan ada atau tidaknya permasalahan.
- c. Untuk mempelajari hubungan atau asosiasi antara faktor-faktor penyebab dengan variabel tujuan.
- d. Untuk mempelajari perbedaan antara kelompok individu secara deskriptif, meliputi nilai-nilai statistik variabel tujuan dan asosiasi antara faktor-penyebab dengan variabel tujuan. (Agung, 2016).

## 2.13 Analisis Sentiment

Analisis sentimen atau opinion mining merupakan perpaduan dari data mining dan *text mining*, suatu teknik untuk menganalisa pendapat, sentimen, evaluasi, sikap, penilaian, perasaan dan emosi seseorang apakah pembicara atau penulis berkenan dengan suatu topik, produk, layanan, organisasi, individu, ataupun kegiatan tertentu. Dengan analisis sentimen kita bisa mengetahui apakah isi teks itu

bersifat positif atau negatif. Sentimen mengacu pada fokus topik tertentu, pernyataan suatu topik mungkin akan berbeda makna dengan pernyataan sama pada subjek berbeda, oleh karena itu, pada beberapa penelitian didahului dengan menentukan elemen dari sebuah produk yang sedang dibicarakan sebelum memulai analisis sentimen (Romadoni et al., 2020).

Analisis sentimen atau opinion mining merupakan proses memahami, mengekstraksi dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif (Mahayani et al., 2020).

#### **2.13.1 Online Review**

*Online Review* merupakan penjelasan dari sebuah tanggapan yang dilontarkan secara daring melalui jaringan komputer atau internet. *Online review* memiliki arti sebagai ringkasan dari pendapat atau tanggapan mengenai berbagai hal seperti buku, film, berita, dan lainnya, *Online review* juga digunakan sebagai alat untuk meninjau atau penilaian sebuah karya yang dibuat oleh orang lain (Fikria, 2018).

*Online Review* dapat diartikan sebagai komentar yang berisikan kritikan, pujian, dan masukan kepada suatu Lembaga pada *platform* tertentu mengenai sebuah karya (Hayuning tyas, 2020).

Dalam penelitian ini, *online review* merupakan ulasan atau komentar terkait evaluasi dan apa yang dirasakan terkait aplikasi, film, buku, dan berbagai hal lainnya yang disampaikan melalui perangkat (komputer, handphone, dll) dan terkoneksi dengan internet pada suatu situs baik milik perusahaan maupun pihak orang ketiga.

### 2.13.2 Web Scraping

*Web Scraping* adalah proses pengambilan sebuah dokumen semi terstruktur dari internet, umumnya berupa halaman web dalam bahasa markup seperti HTML atau XHTML, dan menganalisis dokumen tersebut untuk diambil data tertentu dari halaman tersebut untuk digunakan bagi kepentingan lain. Menurut Turland, 2010 yang dikutip oleh (Kurniasari, 2018) pada tugas akhirnya.

*Web scraping* atau sering disebut sebagai ekstraksi web yaitu proses pengambilan sebuah informasi dokumen dari sebuah web dan menganalisis dokumen tersebut untuk mengambil data dari halaman tersebut untuk digunakan bagi kepentingan yang lain. Proses *scraping* sangat membantu dalam pengambilan sebuah data yang besar dan proses *scraping* juga dapat merubah data tidak terstruktur menjadi terstruktur (Romadoni et al., 2020).

*Web scraping* dapat disimpulkan bahwa sebuah Teknik yang berfokus untuk mendapatkan data dengan cara mengekstraksi informasi dari media internet maupun website. Dalam proses *web scraping* menurut Josi, et al. 2014 dalam penelitiannya yang dikutip dalam penelitian tugas akhir (Kurniasari, 2018) bahwa terdapat beberapa tahapan diantaranya adalah sebagai berikut :

- a. *Create Scraping Template* : Pembuat program mempelajari dokumen HTML dari *website* yang akan diambil informasinya dari tag HTML yang mengapit informasi yang akan diambil.
- b. *Explore Site Navigation* : Pembuat program mempelajari teknik navigasi pada *website* yang akan diambil informasinya untuk ditirukan pada aplikasi *web scraper* yang akan dibuat.
- c. *Automate Navigation and Extraction* : Berdasarkan informasi yang didapatkan dari langkah 1 dan 2 diatas, aplikasi *web scraper* dibuat untuk mengotomatisasi pengambilan informasi dari *website* yang ditentukan; dan
- d. *Extracted Data and Package History* : Informasi yang didapat dari langkah 3 disimpan dalam tabel atau tabel-tabel *database*.

## 2.14 Text Mining

*Text Mining* adalah sebuah proses pengetahuan intensif dimana pengguna berinteraksi dan bekerja dengan sekumpulan dokumen dengan menggunakan beberapa alat analisis . Selain itu *Text mining* atau *text analytics* adalah istilah yang mendeskripsikan sebuah teknologi yang mampu menganalisis data teks semi-terstruktur maupun tidak terstruktur, hal inilah yang membedakannya dengan data mining dimana data mining mengolah data yang sifatnya terstruktur (Khoirul Abbi Rokhman, Berlilana, 2021).

*Text mining* adalah sebuah sistem otomatis yang dilakukan oleh komputer untuk menganalisis sebuah teks dengan tujuan untuk mendapatkan dan menggali sekumpulan informasi yang penting berkualitas dari sekumpulan teks yang terangkum pada sebuah dokumen. Tugas dari prosedur metode ini adalah menemukan kata-kata yang dapat menjelaskan isi dari suatu dokumen agar nantinya dapat dilakukanya analisis lebih lanjut dengan menggunakan metode statistik, klasifikasi dan asosiasi.

*Text mining* seperti yang sudah dijelaskan sebelumnya yaitu memiliki beberapa tahapan dan tiap tahapan tersebut dapat dijelaskan sebagai berikut :

### 1. *Selection Data*

Pada tahap pertama ini yang harus dilakukan adalah melakukan pengumpulan data yang dibutuhkan proses ini disebut juga data awal, yang nantinya data tersebut akan diseleksi dan masuk kedalam tahap kedua .

### 2. *Pre-Processing*

*Text Preprocessing* adalah sebuah proses dalam mempersiapkan dokumen data yang nantinya akan digunakan dan diproses untuk mengubah data yang tidak terstruktur menjadi terstruktur yang berbentuk dokumen atau teks (Harfian, 2021).

Pada tahap kedua ini : setelah mendapatkan data dari tahapan sebelumnya maka akan dilanjutkan kepada tahapan selanjutnta yaitu *pre-processing* dan *cleansing*, yang meliputi juga proses tokenizing,filtering, steaming dan *tagging*. Pada tahapan ini data akan dibersihkan untuk menghilangkan *noise* pada data yang nantinya akan menyebabkan *missing*

*value*. Setelah proses semua itu sudah dilakukan maka data yang dihasilkan dari proses tersebut siap untuk dilanjutkan ke tahap selanjutnya. Berikut adalah tahapan proses dan tahapan *text preprocessing* :

a. *Case Folding*

*Case Folding* merupakan proses pengubahan karakter huruf menjadi seragam sehingga menjadi huruf kecil.

b. *Cleaning*

*Cleaning* merupakan proses penghilangan karakter-karakter untuk mengurangi noise pada proses pengklasifikasian. Kata yang perlu dihilangkan seperti *username*, *mention* dan sebagainya . Adapun karakter kata atau karakter yang akan dihilangkan yaitu karakter atau simbol (@#\$%^&\*()\_+”:{}<>?!~/[]), angka, link url dan lainnya .

c. *Tokenizing*

*Tokenizing* (tokenisasi) adalah sebuah proses memisahkan teks menjadi sebuah kata, angka huruf, tanda baca dan spasi, atau sering disebut sebagai *token*.

d. *Stopwords*

*Stopwords* berfungsi untuk mengklasifikasikan dan menghilangkan kata sambung untuk mengurangi jumlah kata yang harus diproses oleh sistem, seperti contoh dalam Bahasa Indonesia yaitu “yang”, “di”, “ke” dan masih banyak lainnya atau dalam bahas Inggris seperti berikut yaitu, “of”, “is”, “the”, “or” dan masih banyak yang lainnya (Rofiqi et al., 2019)

e. *Stemming*

*Stemming* adalah sebuah proses untuk memproses berbagai variasi morfologi kata kedalam satu bentuk yang baku.

f. *Normalisasi*

Merupakan proses pengembalian kata-kata yang tidak baku kedalam bahasa baku dalam kamus bahasa Indonesia

g. *Filtering*

*Filtering* merupakan tahap untuk mengambil kata-kata penting dari hasil token.

h. *Negation Handling*

Merupakan proses penanganan kata negasi, dengan penggabungan kata negasi dengan kata selanjutnya sehingga menjadi satu kata baru yang bermakna sama (Harfian, 2021).

3. *Feature selection & Extraction*

Pada tahap ketiga ini : pada tahap ini data yang sudah diolah pada tahap selanjutnya akan diolah kembali menggunakan algoritma atau *feature selection* untuk menyeleksi data agar sistem lebih mudah memprosesnya yang nantinya menghasilkan data yang terstruktur. Kemudian hasil data yang didapatkan dari proses seleksi akan dilakukan kembali proses *extraction* (ekstraksi) , proses ini juga salah satu cara agar suatu objek datanya dapat dikenali oleh sistem dan pada tahap ini juga yang akan mencocokkan kepribadian seseorang dengan data yang didapat.

4. *Labelling & Transformed data*

Setelah proses tahapan ketiga selesai maka pada tahap ke empat ini tinggal menentukan kamus/bahasa secara manual (*dictionary manual*) yang sesuai dengan pemilihan kata dan status tersebut akan dimasukkan kedalam *big five personality*.

5. *Interpretation & Evaluation*

Setelah tahap keempat selesai masuk ketahap terakhir yaitu dengan mengkomparasi data yang sudah diolah dengan algoritma *Naïve Bayes Classifier* untuk meningkatkan pengukuran akurasi dari proses text mining yang sudah dilakukan. (M. P. R. Putra & Wardani, 2020)

## **2.15 Klasifikasi**

Klasifikasi adalah proses untuk menemukan model atau fungsi yang dapat menjelaskan mengenai sebuah kelas data maupun sebuah konsep. Tujuan adanya klasifikasi adalah untuk dapat mengetahui sebuah konsep data kelas dari suatu objek yang memiliki label yang tidak diketahui. Pada tahap klasifikasi dibagi menjadi dua yaitu *fase learning* dan *fase test*. Pada *fase learning*, sebagian data yang telah diketahui kelas datanya diumpankan untuk membentuk model perkiraan. Sedangkan pada fase test model yang sudah terbentuk diuji dengan sebagian data

lainnya untuk mengetahui akurasi dari model tersebut. Bila akurasinya mencukupi model ini dapat dipakai untuk prediksi kelas data yang belum diketahui. (Kurniasari, 2018)

Klasifikasi adalah proses pencarian sekumpulan model atau fungsi yang menggambarkan dan membedakan kelas data dengan tujuan agar model tersebut dapat dipergunakan untuk memprediksi kelas dari suatu objek yang belum diketahui kelasnya. Model itu sendiri diperoleh berdasarkan analisis dari data yang sudah diketahui label kelasnya. Dalam pengklasikasian terdapat dua tahap di dalamnya, yaitu tahap pengamatan dan tahap pengujian. Tahap pengamatan merupakan tahap ketika algoritma membangun model klasifikasi dari data latih yang sudah diketahui label kelasnya. Sedangkan tahap pengujian merupakan langkah untuk menerapkan model tersebut pada data uji sehingga kelas yang sesungguhnya dari data uji dapat diketahui. (Mahayani et al., 2020)

## **2.16 Naïve Bayes Classifier**

Metode *Naïve Bayes Classifier* merupakan salah satu metode yang digunakan untuk melakukan pengelompokan suatu opini dengan baik. *Naïve Bayes Classifier* mampu mengelompokkan komentar orang-orang menjadi positif ataupun negatif terhadap suatu produk atau isu yang sedang berkembang dihalayak ramai. Metode *Naïve Bayes Classifier* bermula dari *teorema Bayes* yaitu memprediksi peluang dimasa yang akan datang dan datanya diambil berdasarkan masa yang lalu (M. W. A. Putra et al., 2020).

*Naïve bayes* adalah sebuah metode yang berlandaskan dengan *teorema Bayes*. Oleh sebab itu, metode ini sangat cocok digunakan jika data yang akan diolah sangat banyak karena sangat sederhana, cepat dan memiliki tingkat keakuratan cukup tinggi dalam memproses data. Algoritma *Naïve bayes* ini dibangun menggunakan data latih. Klasifikasi *Naïve bayes* juga dapat menggabungkan pengetahuan sebelumnya dan pengetahuan baru (Gunawan et al., 2017).



### 2.16.1 Evaluasi Performa Model

Model klasifikasi yang dibuat ialah pemetaan dari suatu baris data dengan keluaran sebuah hasil prediksi kelas/target dari data tersebut. Klasifikasi yang memiliki dua kelas sebagai keluarannya disebut dengan klasifikasi biner. Kedua kelas tersebut biasa direpresentasikan dalam  $\{0,1\}$ ,  $\{+1,-1\}$  atau  $\{\text{positive}; \text{negative}\}$ . (Kurniasari, 2018)

Dalam proses evaluasi klasifikasi terdapat empat kemungkinan yang terjadi dari proses pengklasifikasian suatu baris data. Jika data positif dan diprediksi positif akan dihitung sebagai *true positive*, tetapi jika data itu diprediksi negatif maka akan dihitung sebagai *false negative*. Jika data negatif dan diprediksi negatif akan dihitung sebagai *true negative*, tetapi jika data tersebut diprediksi positif maka akan dihitung sebagai *false positif*. Hasil klasifikasi biner pada suatu dataset dapat direpresentasikan dengan matriks 2 x 2 yang disebut *confusion matrix*. (Fawcett, 2006)

**Tabel 2. 2** *Confusion Matrix*

Prediksi	Class	
	Positive	Negative
	Positive	Negative
	True Positive (TP)	False Positive (FP)
	False Negative (FN)	True Negative (TN)

*Confusion matrix* merupakan suatu alat yang berfungsi untuk menganalisis seberapa baik hasil klasifikasi dalam mengenali tuple dari kelas yang berbeda. *Confusion matrix* juga bisa digunakan untuk mencari nilai akurasi (Romadoni et al., 2020). Pada **Tabel 2.2** adalah sebuah gambaran dari *confussion matrix*. *Confussion matrix* memiliki beberapa rumus umum yang dapat digunakan untuk menghitung performa dalam melakukan proses klasifikasi dan hasil dari proses tersebut akan ditampilkan dalam bentuk persentase.

#### 2.16.1.1 Presisi

*Precision* merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif (M. W. A. Putra et al., 2020). *Precision* disebut juga nilai prediktif positif adalah potongan *instance* yang relevan untuk diambil menurut Purbo, 2019 pada tugas akhir (Hayuning tyas, 2020). Untuk mendapatkan nilai Precision dapat dihitung dengan persamaan pada **Gambar 2.4** berikut:

$$Presisi = \frac{TP}{TP+FP}$$

**Gambar 2. 1** Gambar rumus presisi

Di mana,

$TP = True\ Positive$

$FP = False\ Positive$

#### 2.16.1.2 Akurasi

Accuracy merupakan rasio prediksi benar (positif dan negatif) dengan keseluruhan data (M. W. A. Putra et al., 2020). Accuracy merupakan rata – rata aritmatika dari Presisi dan *Inverse* Presisi (oleh Bias) serta rata –rata aritmatika dari *Recall* dan *Inverse Recall*. Menurut Purbo, 2019 pada tugas akhir (Hayuning tyas, 2020). *Accuracy* dapat dihitung dengan rumus pada **Gambar 2.5** berikut:

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN}$$

**Gambar 2. 2** Gambar rumus akurasi

Di mana,

$TN = True\ Negative$

### 2.16.1.3 Recall

*Recall* (sensitifitas) merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif (M. W. A. Putra et al., 2020). *Recall* yang dikenal juga sebagai sensitivitas merupakan potongan contoh revelan untuk diambil. Menurut Purbo, 2019 pada tugas akhir (Hayuning tyas, 2020). Persamaan untuk menghitung *Recall* adalah sebagai berikut:

$$Recall = \frac{TP}{TP+FN}$$

*Gambar 2. 3 Gambar rumus recall*

Di mana,

*FN = False Negative*

### 2.17 Lexicon based

*Lexicon Based* merupakan suatu sumber teks yang memiliki jumlah kata bersentimen positif lebih banyak dari jumlah kata bersentimen negatif, maka teks sumber tersebut diklasikasikan sebagai teks sumber bersentimen positif, begitu pula sebaliknya. Metode *Lexicon Based* digunakan dengan tujuan untuk menentukan sentimen awal dari suatu kalimat berdasarkan jumlah proporsi kata positif atau negatif yang menjadi penyusun suatu kalimat (Mahayani et al., 2020).

*Lexicon Based* merupakan metode umum untuk analisis sentimen dikarenakan metode yang digunakan praktis. Metode yang digunakan *Lexicon Based* menggunakan kamus sebagai sumber bahasa atau leksikal. *Lexicon Based* sebagai pengklasifikasi sebuah sentimen dari setiap opini sehingga sebuah kalimat sentimen dapat dikelas-kelaskan sesuai kelas negatif dan positif. Dengan metode ini sebuah data kalimat sentimen akan lebih mudah untuk diklasifikasi. (Mahendrajaya et al., 2019).

## 2.18 Wordcloud

*Wordcloud* adalah visualisasi dari data teks, digunakan untuk menggambarkan metadata kata kunci (tag), yang biasanya merupakan berbentuk 16 kata tunggal, dengan warna dan ukuran setiap tag yang menunjukkan kadar pentingnya suatu tag (Setyobudi et al., 2018). *Word cloud* berisi kumpulan kata-kata yang membentuk seperti awan kata dengan tampilan visualisasi semakin besar kata diawan, maka kata tersebut frekuensi munculnya semakin banyak.

*Wordcloud* bisa diartikan sebagai visualisasi kata-kata dimana ukurannya disesuaikan dengan frekuensi munculnya kata tersebut. *Wordcloud* akan memudahkan pembaca untuk melihat kata-kata yang sering muncul terkait suatu topik (Inayah & Purba, 2021).

## 2.19 Asosiasi

Istilah Korelasi sering digunakan untuk menyatakan antara hubungan dua atau lebih variabel yang bersifat kuantitatif, sedangkan asosiasi memiliki artian sebagai sebuah keeratan hubungan antara dua atau lebih variabel yang sifatnya kualitatif. Asosiasi merupakan sebuah proses pencarian hubungan antar elemen data. Sering disebut *market basket analysis* dalam dunia retail (Miner et al., 2012). Asosiasi juga salah satu proses pencarian hubungan antara satu elemen klasifikasi yang bersifat positif dengan elemen klasifikasi yang bersifat negatif, sehingga dari hasil proses klasifikasi tersebut dapat menghasilkan sebuah informasi yang nantinya dapat dijadikan referensi dan masukan untuk pihak pengembang maupun pemilik perusahaan aplikasi Dana terkait topik apa saja yang sering dibahas dalam kolom ulasan aplikasi Dana yang dituangkan pengguna dalam situs *google play*. (Kurniasari, 2018). Pada **Gambar 2.12** adalah contoh rumus asosiasi kata.

$$r = \frac{N \sum X_i Y_i - (\sum X_i)(\sum Y_i)}{\sqrt{\{N \sum X_i^2 - (\sum X_i)^2\} \{N \sum Y_i^2 - (\sum Y_i)^2\}}}$$

**Gambar 2. 4** Rumus Asosiasi Kata

## 2.20 Diagram Sebab-Akibat (*Fishbone Diagram*)

Diagram sebab-akibat (*fishbone diagram*) adalah sebuah Teknik grafis yang digunakan untuk mengurutkan dan menghubungkan antara interaksi dan factor-faktor yang berpengaruh lainnya dalam suatu proses. Diagram ini berfungsi untuk mengetahui faktor-faktor apa saja yang berpengaruh kepada kualitas sebuah output, sehingga diharapkan ketika sudah mengetahui faktor masalah yang berpengaruh kepada hasil output yang di inginkan maka dapat langsung diperbaiki pada akar permasalahannya. Menurut (fauziah,2009) yang dikutip oleh (Kurniasari, 2018) pada tugas akhirnya. Diagram sebab-akibat juga memiliki beberapa manfaat diantaranya adalah sebagai berikut :

1. Dapat menggunakan kondisi yang sesungguhnya untuk tujuan perbaikan kualitas produk atau jasa, lebih efisien dalam penggunaan sumber daya, dan dapat mengurangi biaya.
2. Dapat membuat suatu standarisasi operasi yang ada maupun yang direncanakan.
3. Dapat mengurangi dan menghilangkan kondisi yang menyebabkan ketidaksesuaian produk atau jasa, dan keluhan pelanggan.

Langkah-langkah dalam analisis fishbone adalah:

1. Menyiapkan sesi sebab-akibat.
2. Mengidentifikasi akibat.
3. Mengidentifikasi berbagai kategori.
4. Menemukan sebab-sebab potensial dengan cara sumbang saran.
5. Mengkaji kembali setiap kategori sebab utama.
6. Mencapai kesepakatan atas sebab-sebab yang paling mungkin.

## BAB III

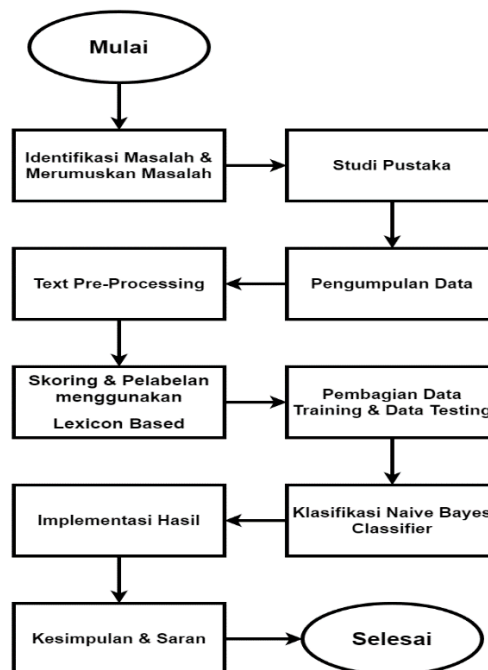
### METODOLOGI PENELITIAN

#### 3.1 Objek Penelitian

Penelitian ini akan dilakukan pada situs resmi *google play* pada ulasan aplikasi *dana* yang dimana peneliti mendapatkan data penelitian secara langsung untuk mengetahui bagaimana proses ekstraksi data sehingga ulasan tersebut dapat bernilai positif maupun negatif.

#### 3.2 Tahapan Kerangka Penelitian

Pada tahap ini akan menampilkan dan memvisualisasikan tahapan kerangka penelitian, pada **Gambar 3.1** berikut adalah visualisasi diagramnya :



*Gambar 3. 1 Tahapan alur Kerangka Penelitian*

Untuk mendeskripsikan **Gambar 3.1** pada tahapan alur kerangka penelitian, maka peneliti akan membuat sub bab yang berisi tentang penjelasan tiap tahapan yang ada pada pada gambar tersebut. Berikut sub bab penjelasanya .

### 3.3.1 Identifikasi dan Merumuskan masalah

#### a. Identifikasi Masalah

Berdasarkan latar belakang yang sudah dijelaskan sebelumnya dapat diidentifikasi masalahnya yaitu cukup banyaknya ulasan yang masuk pada kolom review aplikasi dana tetapi, belum adanya permodelan klasifikasi yang menjelaskan mengenai ulasan review yang bersifat positif, negatif pengguna terhadap aplikasi dana.

#### b. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan sebelumnya, terdapat permasalahan yang akan diangkat dalam skripsi ini yaitu sebagai berikut :

1. Bagaimana membuat permodelan klasifikasi yang dapat menganalisis tiap ulasan yang bersifat positif, negatif, netral yang datanya didapatkan langsung dari situs *google play*.
2. Bagaimana hasil dari penerapan metode *Naïve Bayes Classifier* dalam mengklasifikasikan data ulasan pengguna aplikasi *Dana* menjadi kelas positif dan negatif ?
3. Informasi topik apa yang diperoleh dalam proses klasifikasi dan asosiasi teks yang telah dilakukan ?
4. Bagaimana cara memperbaiki hasil ulasan negatif yang didapat pada saat proses klasifikasi dan asosiasi ?

### 3.3.2 Studi Pustaka

Pada tahap ini peneliti dalam melakukan penelitian melakukan studi Pustaka yang bertujuan sebagai landasan peneliti untuk melakukan penelitian dengan cara membaca berbagai macam jurnal-jurnal, buku yang terkait dan mengumpulkan jurnal penelitian sebelumnya untuk dianalisis agar dapat menentukan metode dan kasus apa yang cocok untuk dijadikan objek penelitian.

### 3.3.3 Pengumpulan Data

Data yang didapatkan dalam penelitian ini adalah data ulasan yang tersedia pada situs resmi *google play* pada kolom ulasan aplikasi *dana*. Seperti yang digambarkan pada **Gambar 3.2**.

1	Column1	Column2	Column3
2	nama	rating	ulasan
3	Pandi bangkleyan	Diberi rating 1 bintang dari 5 bintang	Aplikasi burii!!!!!! Hay,, sobat semua yang belum download aplikasi burii Mendi
4	Triya Rachmadanti	Diberi rating 1 bintang dari 5 bintang	DANA TOLONG RESPON KELUHAN SAYA. KARTU YG DI DAFTAR KAN UNTUK DANA HILANG. DI DALAM APK MASIH ADA SALDO NYA,
5	Francisca Asti	Diberi rating 1 bintang dari 5 bintang	Maaf saya kasi bintang 1 ya.. sebenarnya sih kl bisa setengah seneng saya.. dari 3 hari yg lalu saya tidak bisa akses dana. Padahal :
6	Algi Kemal	Diberi rating 1 bintang dari 5 bintang	Saya cust baru d DANA, dan saya langsung isi saldo dan berniat kirim uang via DANA. Tapi ternyata harus upgrade premium dlu, a
7	Mia Vallenzt	Diberi rating 1 bintang dari 5 bintang	kecewa banget saya transfer saldo pakek bank CIMB niaga 500.000 blum masuk
8	Irpan Apandi	Diberi rating 5 bintang dari 5 bintang	ntap
9	Maulana Ihsan	Diberi rating 5 bintang dari 5 bintang	tingkat kan..
10	kuda gadis	Diberi rating 2 bintang dari 5 bintang	Beli pls ko ga bisa
11	Ade Rohman	Diberi rating 5 bintang dari 5 bintang	mantap kalo ada 20 bintang aku kasih semua simple padat jelas cepat dan membantu
12	Wedont Halawa	Diberi rating 5 bintang dari 5 bintang	Mantap
13	Dedik Nurrahmad	Diberi rating 5 bintang dari 5 bintang	Transaksi jdi lebih mudah
14	Utta Real	Diberi rating 5 bintang dari 5 bintang	Hay,,,dana
15	Risman Saputra	Diberi rating 5 bintang dari 5 bintang	👍
16	Alan Rushnax	Diberi rating 5 bintang dari 5 bintang	Top
17	Vhyndo3 Andiko	Diberi rating 5 bintang dari 5 bintang	Kenapa transfer udah berjam2 statusnya masi menuugu..sedangkan saya butuh cepat..tolong dana gimana caranya ini?
18	Arnov Family	Diberi rating 3 bintang dari 5 bintang	Sangat membantu
19	Muhammad Riski	Diberi rating 1 bintang dari 5 bintang	Apknya sih bagus tapi kok udah ngga bisa narik pulsanya padahal saya mau isi pulsa kok malah stok habis gitu
20	Erwinsyah Putra azha	Diberi rating 5 bintang dari 5 bintang	Saneattlah membantu oembavaran

**Gambar 3. 2** Contoh Dataset Penelitian.

### 3.3.4 Text Pre-processing

Pada tahap *text pre-processing* data yang sudah dikumpulkan selanjutnya dilakukan proses pembersihan data sedemikian rupa sehingga nantinya akan lebih mudah dalam proses pengolahan data, berikut adalah beberapa tahapan *text pre-processing* :

#### a. Case Folding

*Case Folding* merupakan proses pengubahan semua karakter huruf menjadi seragam sehingga menjadi huruf kecil, sedangkan karakter lain selain huruf dihilangkan maka dianggap *delimiter*.

#### b. Cleaning

*Cleaning* merupakan proses penghilangan karakter-karakter untuk mengurangi noise pada proses pengklasifikasian. Kata yang perlu dihilangkan seperti *username*, *mention* dan sebagainya . Adapun karakter kata atau karakter yang akan dihilangkan yaitu karakter atau simbol (@#\$%^&\*()\_+~:}{<>?!~/[], angka, link url dan lainnya .



c. *Tokenizing*

*Tokenizing* (tokenisasi) adalah sebuah proses memisahkan teks menjadi sebuah kata, angka huruf, tanda baca dan spasi, atau sering disebut sebagai *token*.

d. *Stopwords*

*Stopwords* berfungsi untuk mengklasifikasikan dan menghilangkan kata sambung untuk mengurangi jumlah kata yang harus diproses oleh sistem, seperti contoh dalam Bahasa Indonesia yaitu “yang”, “di”, “ke” dan masih banyak lainnya atau dalam bahasa Inggris seperti berikut yaitu, “of”, “is”, “the”, “or” dan masih banyak yang lainnya (Rofiqi et al., 2019)

e. *Filtering*

Filtering merupakan tahap untuk mengambil kata-kata penting dari hasil token.

### 3.3.5 Pelabelan

Pada tahapan pelabelan disini data yang sudah masuk tahap *pre-processing* akan diproses selanjutnya untuk diberikan pelabelan data positif dan negatif, pada tahapan pelabelan tersebut peneliti menggunakan bantuan *Lexicon based* berbahasa Indonesia

### 3.3.6 Klasifikasi Naïve Bayes Classifier

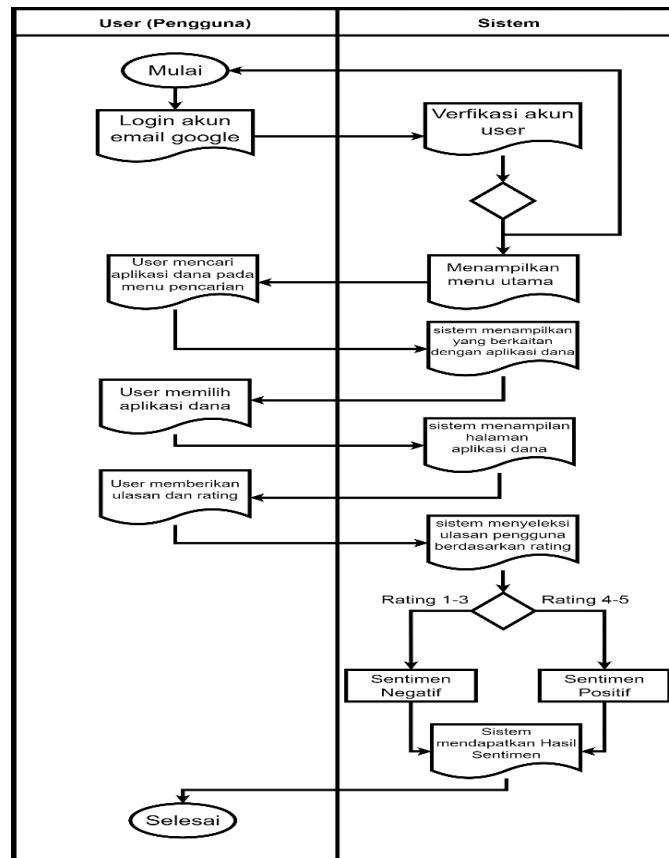
Setelah mendapatkan nilai dan label pada tahapan pelabelan, tahap selanjutnya adalah menguji tingkat akurasi metode *naïve bayes classifier* dalam memproses data dan mendapatkan hasil nilai *confussion matrix*.

### 3.3.7 Hasil

Setelah tahapan pengujian pada klasifikasi selesai maka akan didapatkan hasil output visualisasi kata, nilai akurasi, *confussion matrix*.

### 3.3 Analisis Sistem Berjalan

Pada tahap analisis sistem berjalan ini akan menjelaskan dan menampilkan tahapan sistem berjalan, pada **Gambar 3.3** berikut adalah gambaran visualisasi diagramnya :

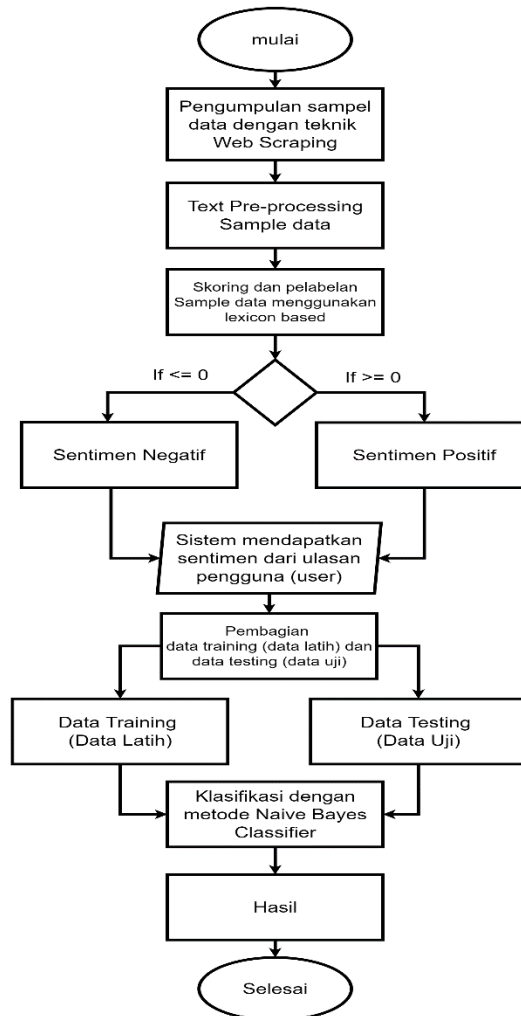


**Gambar 3. 3** Flowmap Sistem Berjalan

Pada **Gambar 3.3** menjelaskan mengenai sistem berjalan dalam menganalisis sentimen pengguna pengguna aplikasi Dana pada *Google play* dengan indikator penilaian (rating) dan pemberian ulasan. Dimulai dari pengguna yang melakukan *login* akun *e-mail* atau *google* lalu sistem memverifikasi akun pengguna, setelah itu sistem menampilkan menu utama, lalu user mencari dan memilih aplikasi dana, selanjutnya sistem menampilkan halaman aplikasi dana, lalu pengguna memberikan rating dan ulasan. Selanjutnya sistem menyeleksi berdasarkan rating untuk menentukan apakah ulasan tersebut bersifat negatif atau positif. Setelah seleksi berdasarkan rating maka didapatkan hasil analisis sentimen tersebut.

### 3.4 Sistem Usulan

Pada tahap sistem usulan ini akan menjelaskan dan menampilkan tahapan sistem yang akan diusulkan, pada **Gambar 3.4** berikut adalah gambaran visualisasi diagramnya :



**Gambar 3. 4** Tampilan Sistem Usulan

Pada **Gambar 3.4** menjelaskan mengenai tahapan sistem yang akan diusulkan diawali dengan tahapan pengumpulan data dengan menggunakan web scraping kemudian data tersebut memasuki proses text-preprocessing untuk memasuki tahapan pembersihan data. Setelah itu memasuki tahapan perhitungan nilai tiap ulasan yang bertujuan untuk mengetahui hasil label sentimen tiap ulasan yang diberikan oleh pengguna. Selanjutnya melakukan pembagian data menjadi data training dan data testing, pada penelitian ini peneliti melakukan 2 (dua) proses

perbandingan data latih (*data training*) dan data uji (*data testing*) sebesar 80%:20% dan 70%:30% , kemudian dilakukan proses tahapan klasifikasi menggunakan metode naïve bayes classifier untuk mendapatkan hasil nilai *confussion matrix*, nilai akurasi, nilai presisi, nilai *recall*.

### **3.5 Populasi dan Sampel Penelitian**

Populasi yang digunakan dalam penelitian ini adalah data ulasan/review pada aplikasi Dana yang terdapat pada situs resmi *google play*, yang diambil periode 20 April 2021 hingga 25 Mei 2021. Sedangkan sampel yang digunakan dalam penelitian ini adalah data ulasan pada aplikasi dana yang berbahasa Indonesia. Total data sampel yang akan diolah sebanyak 10001 data ulasan berbahasa Indonesia .

### **3.6 Jenis dan sumber Data Penelitian**

Jenis data yang akan digunakan dalam proses penelitian ini adalah data primer, data yang didapatkan dengan cara langsung pada beberapa ulasan dari pengguna aplikasi dana dengan menggunakan teknik *web scraping* dari halaman situs *google play* pada aplikasi dana dengan alamat *URL* situs sebagai berikut <https://play.google.com/store/apps/details?id=id.dana&showAllReviews=true>. menggunakan ekstensi aplikasi *scraping* bawaan dari browser *google chrome* yaitu *data miner*.

### **3.7 Metode pengumpulan Data**

Pada tahap ini, peneliti akan mengumpulkan data dari situs resmi *google play* yang memuat berbagai macam ulasan dan komentar yang ada pada kolom ulasan aplikasi dompet digital dana.

Metode pengumpulan data yang digunakan dalam penelitian kali ini adalah dengan cara melakukan teknik *Web scraping* dari beberapa ulasan (review) aplikasi dompet digital dana yang ada pada situs resmi *google play* lalu didapatkan beberapa data ulasan berbahasa Indonesia.

Teknik *Web scraping* merupakan proses pengambilan dokumen data yang semi terstruktur dari internet yang berisi beberapa data dokumen yang digunakan untuk kepentingan tertentu dan berfokus pada mengekstraksi data dalam jumlah yang bervariasi. (Astuti, 2020)

Alat atau instrumen yang digunakan dalam pengambilan atau pengumpulan data pada penelitian ini adalah dengan menggunakan bantuan aplikasi ekstensi pada google chrome yaitu aplikasi *data miner*, setelah melakukan proses scraping pada situs google play maka akan didapatkan data ulasan dalam bentuk file *CSV* yang bisa diakses dan dibaca dengan menggunakan *Microsoft excel*.

### 3.8 Variabel Penelitian

Variabel yang akan digunakan dalam proses penelitian kali ini yaitu sebagai berikut :

***Tabel 3. 1 Variabel Penelitian***

Variabel	Definisi Variabel
Nama ( <i>user</i> )	Nama pengguna ( <i>user</i> ) yang melakukan ulasan mengenai aplikasi dana pada situs <i>google play</i> .
Rating	Penilaian pengguna mengenai aplikasi dana
Ulasan ( <i>review</i> )	komentar dari pengguna pada aplikasi Dana yang ada di dalam situs Google play

### 3.9 Metode analisis Data

Metode atau teknik analisis yang akan digunakan dalam penelitian ini adalah sebagai berikut :

#### a. Metode Teknik *Web Scraping*

Metode Teknik *Web Scraping* merupakan sebuah metode yang digunakan untuk mendapatkan atau mengumpulkan data ulasan aplikasi dana secara online yang bersumber dari situs *Google Play*

b. Metode analisis deskriptif

Metode analisis deskriptif merupakan metode yang digunakan untuk menggambarkan secara umum mengenai ulasan aplikasi dana pada kolom review pada situs google play

c. Metode *Machine Learning* .

Metode *Machine Learning* yang akan digunakan pada penelitian ini adalah metode *Naïve Bayes Classifier*. Metode tersebut digunakan untuk mengklasifikasikan tiap ulasan-ulasan pengguna menjadi beberapa kelompok seperti kelompok kelas positif, kelas negatif dan kelas netral.

d. Metode *Association*.

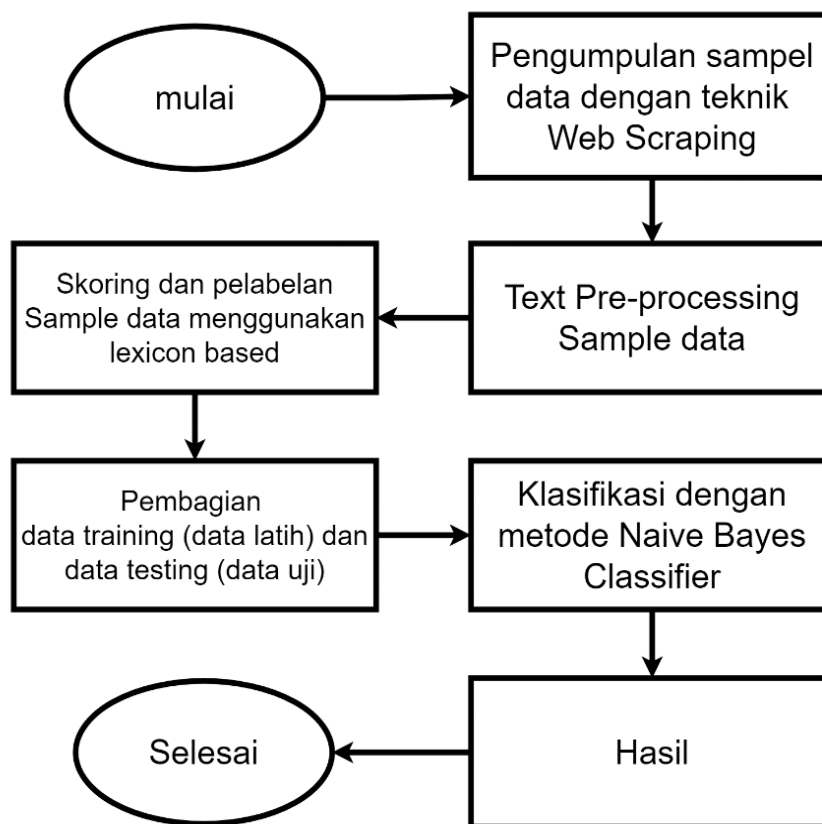
Metode *Association* merupakan sebuah metode yang digunakan untuk mengidentifikasi dan membentuk pola kata yang saling berasosiasi dengan satu kata dengan kata lainnya, guna untuk mendapatkan informasi yang dianggap penting.

## BAB IV

### PERANCANGAN SISTEM DAN IMPLEMENTASI

#### 4.1 Alur Implementasi

Pada tahap ini akan membahas dan menampilkan mengenai tahapan alur sistem, pada **Gambar 4.1** berikut adalah gambaran visualisasi diagramnya :



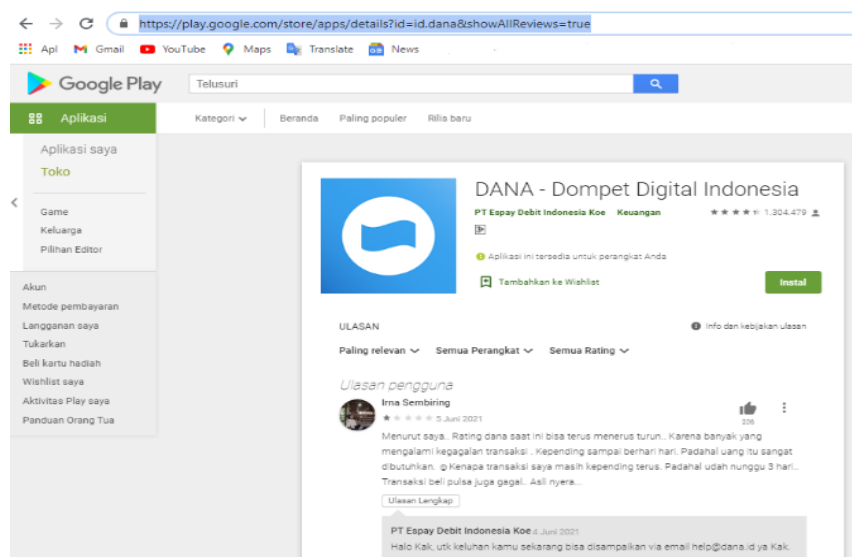
**Gambar 4. 1** Tahapan Alur Implementasi

Untuk mendeskripsikan **Gambar 3.1** Menjelaskan mengenai alur implementasi diawali dengan tahapan pengumpulan data dengan menggunakan web scraping kemudian data tersebut memasuki tahapan proses *text-preprocessing* untuk memasuki tahapan pembersihan data. Setelah itu memasuki tahapan

perhitungan nilai tiap ulasan yang bertujuan untuk mengetahui hasil label sentimen tiap ulasan yang diberikan oleh pengguna. Selanjutnya melakukan pembagian data menjadi data training dan data testing, kemudian dilakukan proses tahapan klasifikasi menggunakan metode naïve bayes classifier untuk mendapatkan hasil nilai *confussion matrix*, *nilai akurasi*, *nilai presisi*, *nilai recall*.

## 4.2 Pengumpulan Data

Pengambilan sampel data ulasan yang akan diproses pada penelitian tugas akhir ini bersumber dari situs resmi website google play pada kolom ulasan aplikasi dana yang beralamat URL pada **Gambar 4.2** (<https://play.google.com/store/apps/details?id=id.dana&showAllReviews=true>). Pada proses pengumpulan data ini variabel data yang diambil adalah nama pengguna, rating serta ulasan pengguna mengenai aplikasi dana.

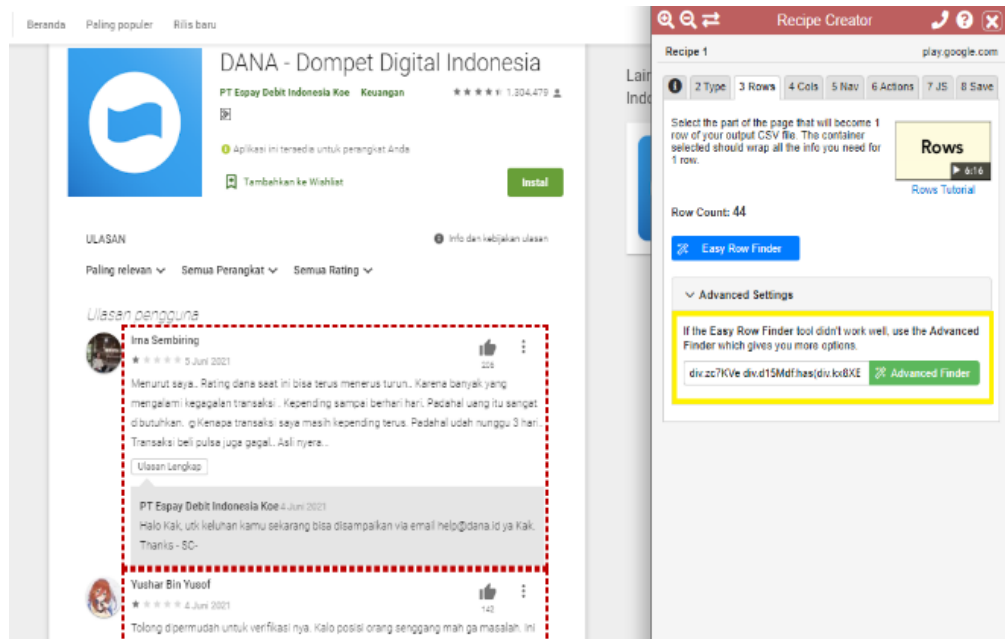


**Gambar 4. 2** Alamat situs Googleplay yang akan discraping

Xpath (*The XML Path language*) adalah sebuah Bahasa maupun script yang berfungsi dan dapat digunakan untuk mendapatkan sebuah informasi dalam sebuah dokumen XML. Sedangkan XML (*Extensible Markup Language*) sendiri adalah sebuah Bahasa markup yang menyediakan kata-kata dan tag yang dapat menggambarkan dokumen dan dapat mengidentifikasi tiap bagian-bagiannya.



Pada penelitian ini Xpath digunakan untuk mempermudah dalam menyeleksi kolom maupun baris mana saja yang ingin discraping , dalam hal ini baris yang ingin discraping adalah pada **Gambar 4.3** dengan Xpath (Ditandai dengan garis warna kuning) ( `div.zc7KVe div.d15Mdf:has(div.kx8XBd)` ) .



**Gambar 4.3** Baris selector Xpath dalam proses scraping

Setelah menyeleksi baris yang diinginkan untuk dilakukan proses scraping. Selanjutnya untuk memperjelas variabel mana saja yang akan diteliti maka diperlukanya menyeleksi kolom yang mana saja yang akan dijadikan variabel, sedangkan pada penelitian ini variabel yang akan diteliti terdapat 3 yaitu sebagai berikut pada **Gambar 4.4** (ditandai dengan garis hijau) :

### 1. Variabel nama pengguna (user)

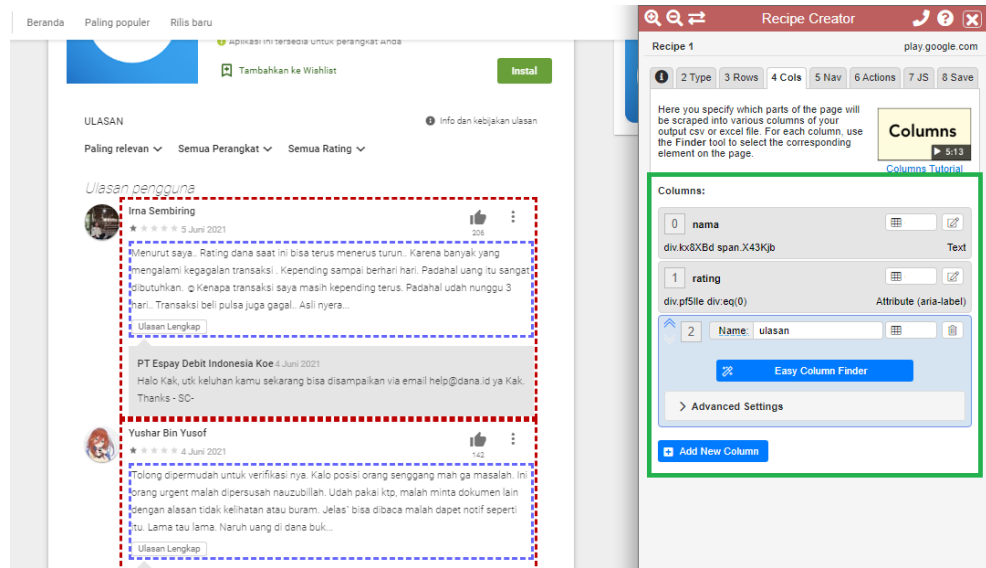
*Selector Xpath* yang akan diseleksi adalah (`div.kx8XBd span.X43Kjb`). Dengan attribut yang akan diseleksi adalah dalam bentuk *text*.

### 2. Variabel rating (nilai)

*Selector Xpath* yang akan diseleksi adalah (`div.pf51Ie div:aria-label`). Dengan attribut yang akan diseleksi adalah *other attribute* : `*aria-label`.

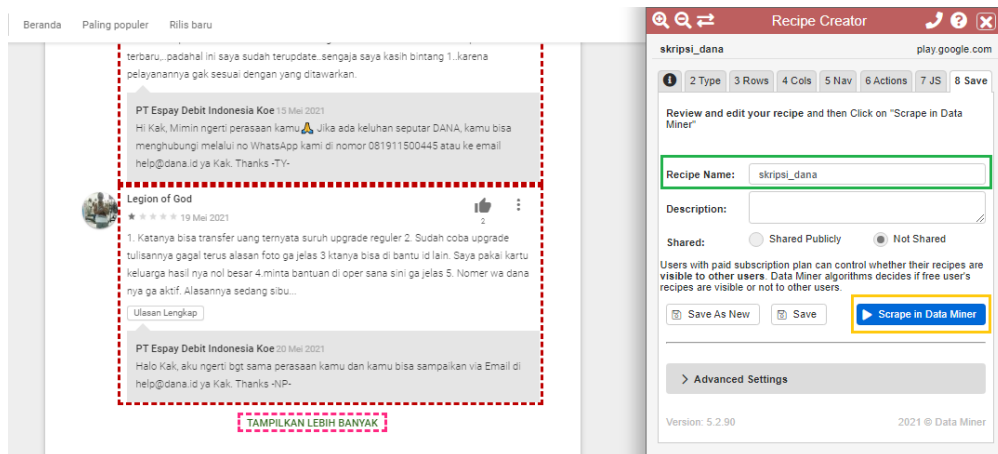
### 3. Variabel Ulasan (review)

*Selector Xpath* yang akan diseleksi adalah (`div.UD7Dzf`). Dengan attribut yang akan diseleksi adalah *text*.



**Gambar 4. 4** Kolom selector *Xpath* dalam proses *scraping*

Lalu, setelah semua *selector Xpath* yang ingin diseleksi sudah terdata pada aplikasi ekstension *Data miner*. Selanjutnya *disave* masukan nama *file* (pada gambar 4.5 ditandai dengan warna hijau) yang diinginkan setelah proses *scraping*, lalu masuk pada tahapan *scraping* (seperti pada gambar 4.5 ditandai warna kuning) untuk mendapatkan data untuk diteliti, seperti pada **Gambar 4.5**.



**Gambar 4. 5** Proses *scraping* data

Setelah proses *Scraping* selesai, selanjutnya file dapat diunduh dalam bentuk file CSV untuk masuk kedalam tahapan selanjutnya pada proses penelitian ini, yaitu proses *text pre-processing* untuk membersihkan data tersebut, berikut pada **Tabel4.1** adalah tampilan data hasil *Scraping* .

**Tabel 4. 1** Tampilan contoh data hasil proses Scraping

No	Nama	Rating	Ulasan
1	Pandi bangkleyan	1	Aplikasi buriiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii!!!!!! Hay,, sobat semua yang belum download aplikasi buriiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiiii Mending di batalin aja deh Aplikasi nya nggsk bisa di pakai Nyesel banget saya mendownload nya!!!!!!!!!!!!!!
2	Triya Rachmadanti	1	DANA TOLONG RESPON KELUHAN SAYA. KARTU YG DI DAFTAR KAN UNTUK DANA HILANG. DI DALAM APK MASIH ADA SALDO NYA. TOLONG RESPON!!!! BALIKIN SALDO SAYA. BARU HAPUS AKUN. TOLONG RESPON. SAYA BERUSAHA UNTUK Mencari JALAN KELUAR MALAH DARI DANA NYA SURUH KE WA. GILIRAN KE WA MALAH SIBUK!!!!!!!!!!!!!!
3	Francisca Asti	1	Maaf saya kasi bintang 1 ya.. sebenarnya sih kl bisa setengah seneng saya.. dari 3 hari yg lalu saya tidak bisa akses dana. Padahal sudah update, masih tetap tidak bisa. Masa selama 3 hari tidak ada perbaikan.. baca komen2 yg lain kok senada meski masalahnya beda..
4	Algi Kemal	1	Saya cust baru d DANA, dan saya langsung isi saldo dan berniat kirim uang via DANA. Tapi ternyata harus upgrade premium dlu, alhasil saya mencoba upgrade ke premium. Tapi kok rasanya sulit sekali cuma upgrade premium. Ini udah ke 4x nya saya mencoba upgrade dan gagal

			terus dengan alasan "foto KTP ti...Ulasan Lengkap
5	Mia Vallenzt	1	kecewa banget saya transfer saldo pakek bank CIMB niaga 500.000 blum masuk
6	Irpan Apandi	5	ntap
7	Maulana Ihsan	5	tingkat kan..
8	kuda gadis	2	Beli pls ko ga bisa
9	Ade Rohman	5	mantap kalo ada 20 bintang aku kasih semua simple padat jelas cepat dan membantu
10	Wedont Halawa	5	Mantap

Sekumpulan data hasil *scraping* tersebut selanjutnya akan dilakukan proses penghapusan kolom indikator *variabel* yang tidak digunakan dalam proses pengolahan data. Kolom indikator *variabel* yang dihapus diantaranya adalah kolom nama pengguna serta rating, seperti yang akan ditampilkan pada **Tabel 4.2** Berikut:

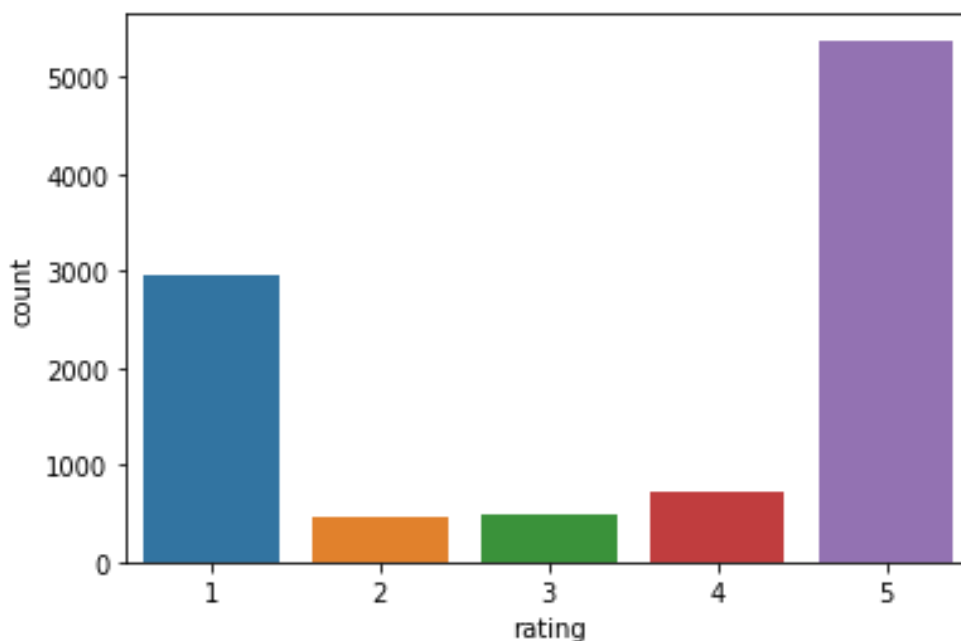
***Tabel 4. 2 Hasil Penghapusan Kolom***

No	Ulasan
1	Aplikasi burii!!!!!!!!!!!!!! Hay,, sobat semua yang belum download aplikasi burii Mending di batalin aja deh Aplikasi nya nggsk bisa di pakai Nyesel banget saya mendownload nya!!!!!!!!!!!!!!
2	DANA TOLONG RESPON KELUHAN SAYA. KARTU YG DI DAFTAR KAN UNTUK DANA HILANG. DI DALAM APK MASIH ADA SALDO NYA. TOLONG RESPON!!!! BALIKIN SALDO SAYA. BARU HAPUS AKUN. TOLONG RESPON. SAYA BERUSAHA

	UNTUK Mencari jalan keluar malah dari Dana nya suruh ke WA. Giliran ke WA malah sibuk!!!!!!!!!!!!!!
3	Maaf saya kasi bintang 1 ya.. sebenarnya sih kl bisa setengah seneng saya.. dari 3 hari yg lalu saya tidak bisa akses dana. Padahal sudah update, masih tetap tidak bisa. Masa selama 3 hari tidak ada perbaikan.. baca komen2 yg lain kok senada meski masalahnya beda..
4	Saya cust baru d DANA, dan saya langsung isi saldo dan berniat kirim uang via DANA. Tapi ternyata harus upgrade premium dlu, alhasil saya mencoba upgrade ke premium. Tapi kok rasanya sulit sekali cuma upgrade premium. Ini udah ke 4x nya saya mencoba upgrade dan gagal terus dengan alasan "foto KTP ti...Ulasan Lengkap
5	kecewa banget saya transfer saldo pakek bank CIMB niaga 500.000 blum masuk
6	ntap
7	tingkat kan..
8	Beli pls ko ga bisa
9	mantap kalo ada 20 bintang aku kasih semua simple padat jelas cepat dan membantu
10	Mantap

#### 4.3 Analisis Deskriptif

Analisis deskriptif dalam penelitian ini bertujuan untuk mendeskripsikan dan menggambarkan data ulasan tersebut secara umum serta jumlah total mengenai pemberian rating pengguna aplikasi dana berdasarkan pada data ulasan yang ada pada situs *google play*. Serperti yang ditampilkan pada **Gambar 4.6** Berikut :



**Gambar 4. 6** Diagram Rating Ulasan Berbahasa Indonesia

Berdasarkan hasil yang ditampilkan pada diagram **Gambar 4.6** di atas bahwa rata-rata pengguna aplikasi *dana* memberikan rating/penilaian yang baik, dari total data ulasan sebanyak 10.001 data dengan rincian diantaranya adalah sebagai berikut :

- Penilaian dengan rating 5 sebanyak 5.376 ulasan
- Penilaian dengan rating 4 sebanyak 716 ulasan
- Penilaian dengan rating 3 sebanyak 489 ulasan
- Penilaian dengan rating 2 sebanyak 472 ulasan
- Penilaian dengan rating 1 sebanyak 2.947 ulasan

#### **4.4 Text-Preprocessing.**

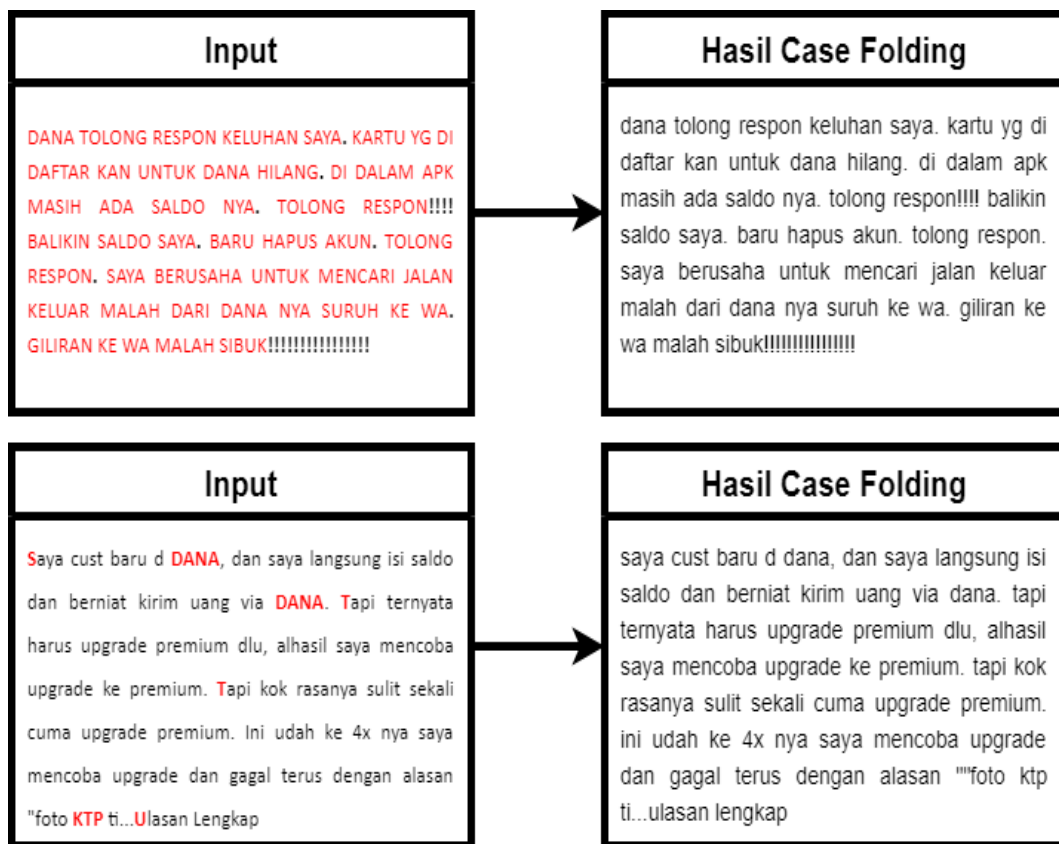
Data ulasan aplikasi *dana* yang didapatkan dari hasil dari *scraping* pada situs *google play* belum bisa langsung digunakan dikarenakan data tersebut masih berbentuk sekumpulan teks yang tidak beraturan /terstruktur dimana data tersebut masih banyak *noise* sehingga informasi yang ada pada tersebut akan sulit diekstrak secara langsung. Oleh sebab itu, maka data tersebut harus diseleksi dan dibersihkan terlebih dahulu agar lebih seragam dan terstruktur sehingga lebih mudah dalam proses Analisa dan ekstrasi sebuah data informasi yang ada pada data tersebut.

Pada tahap preprocessing, akan dilakukan pembersihan data menggunakan metode text mining. Beberapa tahap yang akan dilakukan diantaranya adalah sebagai berikut :

- |  |   |
|--|---|
| <ul style="list-style-type: none"> <li>a. <i>Case folding.</i></li> <li>b. <i>Cleaning.</i></li> <li>c. <i>Tokenizing</i></li> </ul> | <ul style="list-style-type: none"> <li>d. <i>Spelling Normalization</i></li> <li>e. <i>Filtering</i></li> <li>- <i>Stopwords</i></li> </ul> |
|--|---|

#### 4.4.1 Case Folding

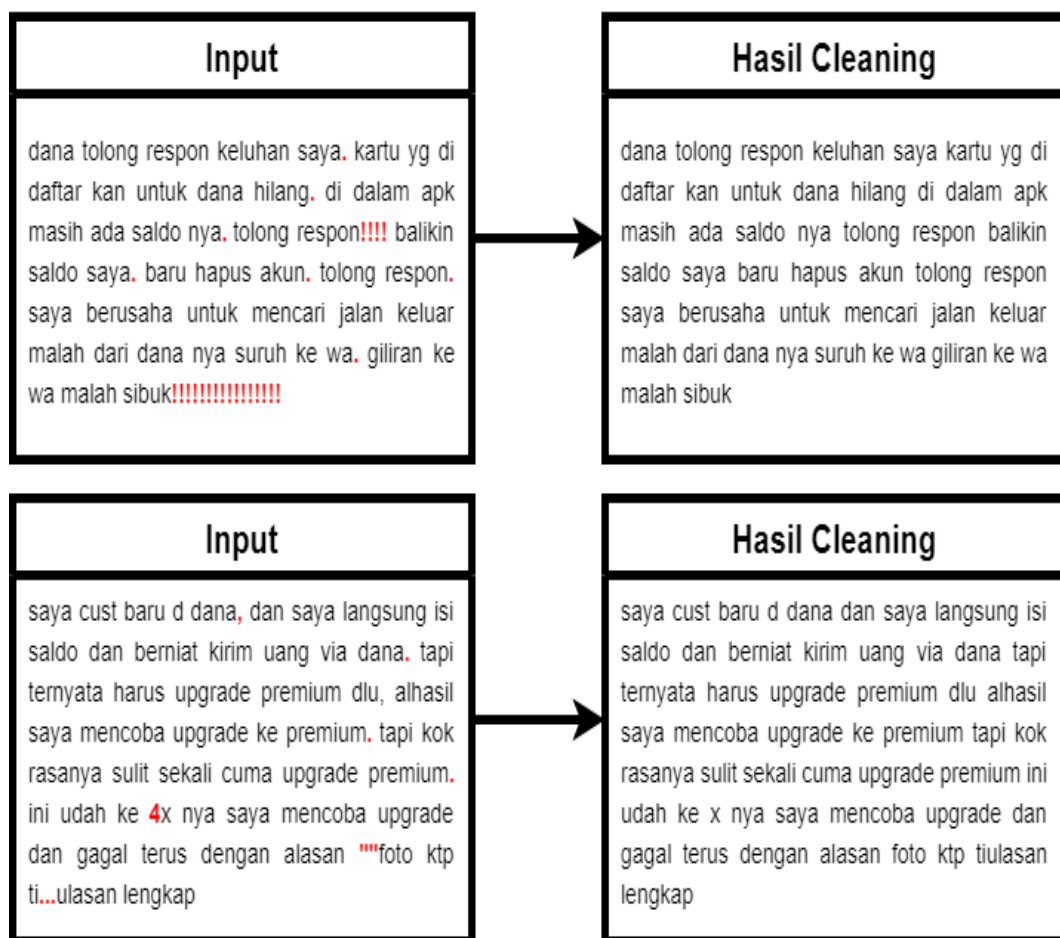
*Case Folding* merupakan proses pengubahan karakter huruf menjadi seragam sehingga menjadi huruf kecil semua. Selain karakter huruf ‘a’ sampai dengan karakter huruf ‘z’ maka simbol dan karakter lain tidak akan diubah, berikut adalah contoh implementasi penggunaan proses *case folding* pada data ulasan (*review*) aplikasi *dana* pada *google play*, seperti pada **Gambar 4.7** berikut:



*Gambar 4. 7 Hasil Case Folding*

#### 4.4.2 Cleaning

Setelah masuk tahapan proses *case folding*, tahap selanjutnya yaitu *Cleaning*. *Cleaning* merupakan proses penghilangan karakter-karakter untuk mengurangi noise pada proses pengklasifikasian. Kata yang perlu dihilangkan seperti *username*, *mention* dan sebagainya . Adapun karakter kata atau karakter yang akan dihilangkan yaitu karakter atau simbol (@#\$%^&\*()\_+”:{}<>?!~/[]), angka, link url ,spasi dan lainnya . seperti contoh pada **Gambar 4.8** berikut ini :

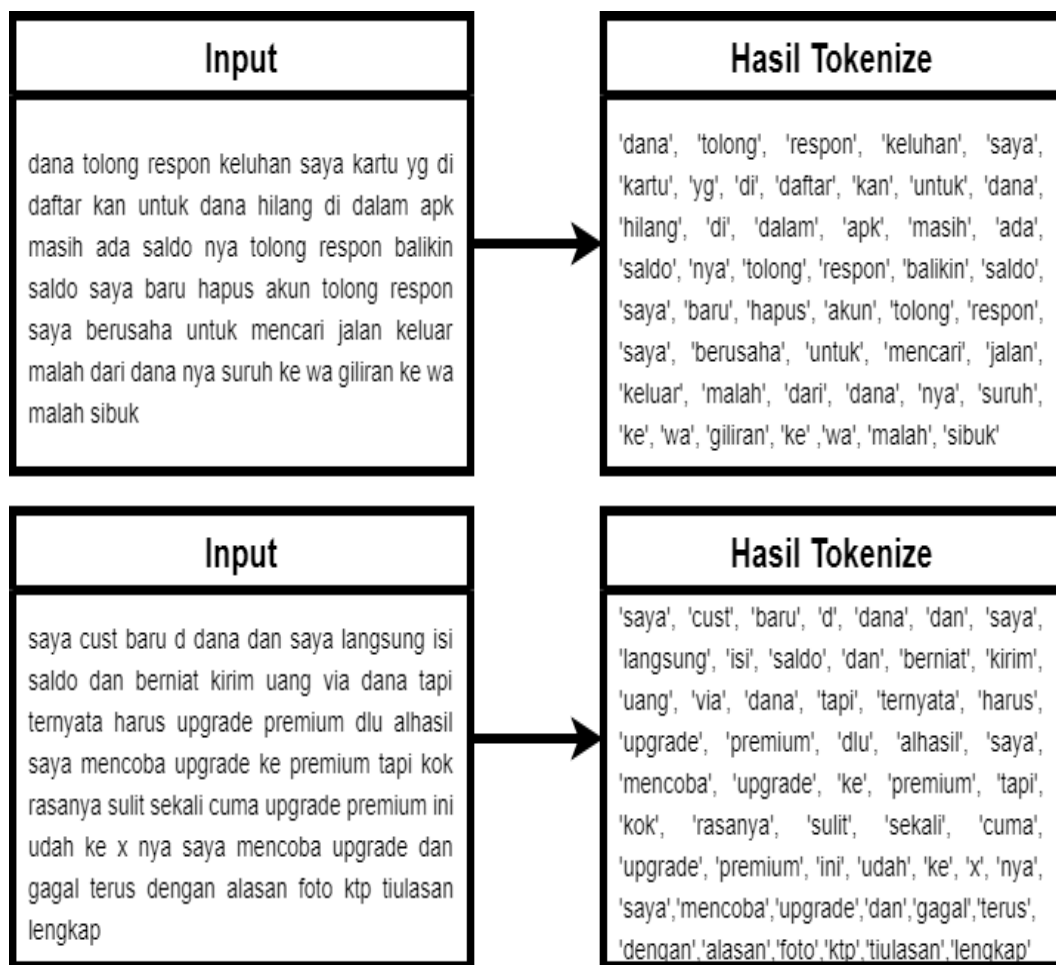


**Gambar 4. 8 Hasil Cleaning**



#### 4.4.3 Tokenizing

Setelah masuk tahapan proses *cleaning*, tahap selanjutnya yaitu *tokenize*. *Tokenizing* (tokenisasi) adalah sebuah proses memisahkan teks menjadi sebuah kata, angka huruf, tanda baca dan spasi, atau sering disebut sebagai *token*. Proses *tokenize* dilakukan dengan tujuan untuk mendapatkan token atau potongan kata yang akan menjadi entitas yang memiliki nilai dalam penyusunan matriks dokumen pada proses selanjutnya. seperti contoh pada **Tabel 4.9** berikut ini :

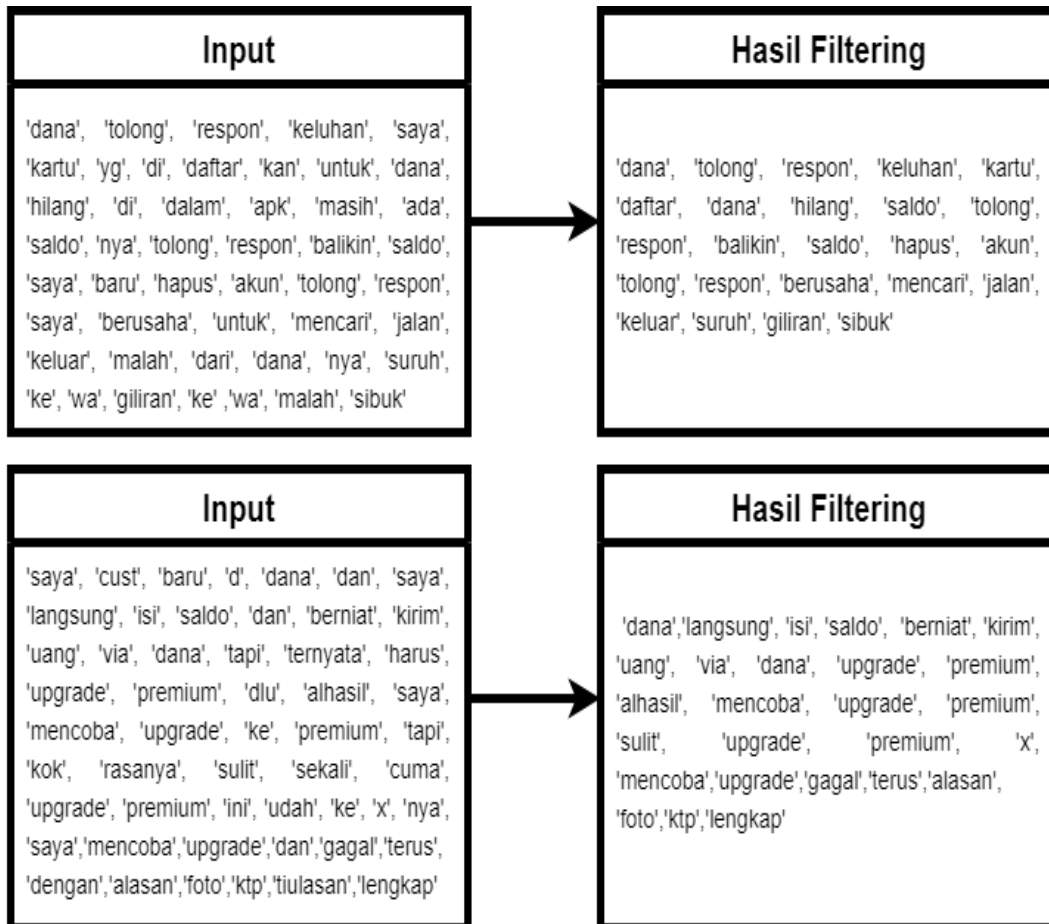


**Gambar 4. 9** Hasil Tokenisasi

#### 4.4.4 Filtering

*Filtering* adalah tahapan *text pre-processing* yang bertujuan untuk pembersihan teks dari stopword. *Stopword* adalah kata-kata yang tidak memiliki makna. Pada proses filtering untuk menghapus stopwords digunakan *library* NLTK,

dimana dalam *library* tersebut hanya terdapat 778 kata *stopword*. *Stopword* tambahan dapat ditambahkan secara manual perkata, penambahan *stopword* juga dapat dibuat terlebih dahulu kamus *stopword* dalam *file csv*. Berikut adalah visualisasi tampilan hasil *filtering* pada **Tabel 4.10** berikut ini:



**Gambar 4. 10 Hasil Filtering**

#### 4.4.5 Spelling Normalization

*Spelling normalization*, merupakan sebuah tahapan *text pre-processing* dengan tujuan untuk memperbaiki kata-kata yang salah penulisanya (*typo*) atau disingkat dengan bentuk tertentu. Misalnya kata “tidak” memiliki banyak bentuk penulisan seperti tdk, gak, nggak, enggak, dan banyak lainnya. *Spelling normalization* juga dapat menyeragamkan kata yang memiliki makna yang sama tetapi penulisanya berbeda, berikut adalah daftar *spelling normalization* seperti contoh pada **Gambar 4.11**.

Spelling Normalization			
knpa	=	kenapa	
knapa	=	kenapa	
mantapp	=	mantap	
mntap	=	mantap	
mantab	=	mantap	
baguss	=	bagus	
bgs	=	bagus	
aja	=	sajah	
ajh	=	sajah	
sy	=	saya	
sya	=	saya	
dlu	=	dulu	
apk	=	aplikasi	
apl	=	aplikasi	
nggk	=	tidak	
g	=	tidak	
udh	=	sudah	
tidakk	=	tidak	
tdk	=	tidak	
sayah	=	saya	
trus	=	terus	
trs	=	terus	
sayaa	=	saya	
app	=	aplikasi	
nyesel	=	menyesal	
blm	=	belum	
kasi	=	kasih	
ksih	=	kasih	
tpi	=	tapi	
blum	=	belum	
blom	=	belum	
pdhl	=	padahal	
slalu	=	selalu	
syaariah	=	syariah	
dln	=	dalam	
mna	=	mana	
downlod	=	unduh	
tlp	=	telepon	
masi	=	masih	
dpt	=	dapat	
verifikasihh	=	verifikasi	

**Gambar 4. 11** Kata Spelling Normalization

#### 4.5 Pelabelan data

Pelabelan data adalah sebuah proses pemberian label untuk membedakan keterangan ulasan yang bersifat positif maupu negatif yang dilakukan dengan cara otomatis dengan bantuan kamus *Lexicon based* untuk mendapatkan parameter indikator kata yang bernilai positif maupun negatif yang berformat file .txt . Berikut adalah rincian jumlah kamus *lexicon* yang digunakan .:

- a. Kamus *lexicon* Positif berbahasa Indonesia : 1182 kata

**Tabel 4. 3** Contoh kata Positif pada kamus *lexicon*

No	Kata Positif	No	Kata Positif
1	Aman	6	Asli
2	Amanah	7	Bangga
3	Ajaib	8	Bagus
4	Alhamdulillah	9	Baik
5	Ambisius	10	Mudah

- b. Kamus *lexicon* Positif berbahasa Indonesia : 2402 kata

**Tabel 4. 4** Contoh kata Negatif pada kamus *lexicon*

No	Kata Negatif	No	Kata Negatif
1	Absurd	6	Agresif
2	Acak	7	Aneh
3	Aib	8	Anjing
4	ambigu	9	Apatis
5	alergi	10	alergi

#### 4.5.1 Skoring

Pada tahapan skoring ini data kamus yang sudah diimport lalu dihitung perkata dengan nilai skor lalu dihitung total nilainya dalam satu kalimat, jika nilai skor kata mendapatkan nilai 1 (satu) maka kata tersebut bernilai ‘positif’ sedangkan jika kata tersebut mendapatkan nilai -1 (satu) maka kata tersebut bernilai negatif. Lalu setelah mendapatkan total skor kata dalam satu kalimat maka nilainya dijumlahkan. Jika hasil total nilai skor dalam satu kalimat  $\geq 0$  maka kalimat

tersebut bernilai “positif”. Sedangkan sebaliknya jika hasil total nilai skor dalam satu kalimat  $\leq 0$  maka kalimat tersebut bernilai “negatif”. Berikut adalah contoh perhitungan nilai skor pada satu kalimat :

**Tabel 4. 5** Perhitungan skoring

Ulasan	Positif	Negatif
maaf kasih bintang lemot	-	maaf
banget pas buka apk isi	-	lemot
pulsa gak		
Total Jumlah Kata	0	2

Sehingga didapatkan hasil perhitungan sebagai berikut :

Skoring = (Jumlah kata positif – Jumlah kata negatif)

Skoring = (0-2)

Skoring = -2 (negatif)

Nilai akhir skoring yang dihasilkan dari simulasi pada **Tabel 4.5** bernilai (-2) atau  $< 0$ , maka hasil klasifikasi yang menjelaskan kalimat tersebut bernilai negatif. Sedangkan untuk total rincian jumlah kelas sentiment positif dan negatif pada penelitian ini akan dijelaskan pada **Tabel 4.6** berikut.

**Tabel 4. 6** Rincian Jumlah total data sentimen

Kelas Sentimen	Jumlah ulasan
Positif	1383
Negatif	8618
Total data	10.001

Berdasarkan penjelasan pada **Tabel 4.6**, Maka disimpulkan bahwa data hasil pelabelan sentimen negatif memiliki frekuensi lebih tinggi dibandingkan sentimen

ulasan positif dari total data sampel keseluruhan sebanyak 10.001 data ulasan. Berikut adalah gambar jumlah detail ulasan sentimen positif maupun negatif secara keseluruhan yang diolah menggunakan tools Rstudio.

The figure shows two panels. The left panel is a screenshot of an R console session where a data frame named 'df' is created with two columns, 'Negatif' and 'Positif', each containing 1000 random values. The right panel is a summary of the data frame 'df', showing it has 2 columns and 2000 rows (1000 for each category).

```
R Console Output:
> df = data.frame(
+   Negatif = c(1000 * runif(1000) - 500, 1000 * runif(1000) - 500),
+   Positif = c(1000 * runif(1000), 1000 * runif(1000))
+ )
> df
      Negatif      Positif
1  -100.42345  100.42345
2   -99.87654  100.12345
3   -99.12345  100.98765
...
1000 -100.12345 100.12345
```

data.frame  
2 x 2

Description: df [2 x 2]

class <fctr>	n <int>
Negatif	8618
Positif	1383

**Gambar 4. 12** Detail jumlah sentimen pada Rstudio

Keseluruhan kata berdasarkan sentimen tersebut sehingga dapat divisualisasikan menggunakan wordcloud, seperti pada **Gambar 4.13** berikut.



**Gambar 4. 13** Wordcloud seluruh sentimen kata

Berikut adalah detail jumlah kata pada *wordcloud* keseluruhan pada sentimen ulasan yang akan digambarkan pada Gambar 4.14 Berikut :

word <chr>	freq <dbl>
dana	2411
aplikasi	1362
bagus	1229
sibuk	775
masuk	754
gak	748
tolong	683
apk	660
saldo	624
membantu	596
sistem	587
banget	559
udah	520
akun	515
uang	501

**Gambar 4. 14** detail jumlah kata pada wordcloud keseluruhan

#### 4.6 Data Latih dan Data Uji

Data latih dalam penelitian ini berfungsi untuk dijadikan sebagai indikator sumber pengetahuan bagi algoritma dalam mengklasifikasikan data maupun untuk memprediksi kelas data baru yang belum pernah ada, sedangkan semakin banyak data latih yang digunakan dalam proses klasifikasi maka hasil yang akan didapat dan diproses oleh sistem semakin baik karena sistem dalam hal ini akan semakin memahami pola dalam sebuah data. Sedangkan data uji digunakan untuk mengukur sejauh mana sistem algoritma dalam memproses dan mengklasifikasikan sebuah data. Dalam penelitian ini peneliti menggunakan beberapa perbandingan data uji latih, pada Tabel 4.7 berikut adalah rincian jumlah perbandingannya.

**Tabel 4. 7** Perbandingan Data Latih dan Data Uji

Sentimen	Jumlah	Data Latih (80%)	Data Uji (20%)	Data Latih (70%)	Data Uji (30%)
Positif	1.383	1.106,4 ~ 1.106	276,6~277	968,1~968	414,9~415
Negatif	8.618	6.894,4 ~ 6.894	1.723,6~1.724	6.032,6~6.033	2.585,4~2.585
Total	10.001	8.000	2001	7001	3.000

Berdasarkan **Tabel 4.8** dijelaskan bahwa perbandingan antara data latih dengan data uji pada penelitian ini adalah sebesar 80% : 20% dan 70% : 30% data ulasan yang berbahasa Indonesia.

#### 4.7 Klasifikasi dengan metode *naïve bayes classifier*

Penelitian ini menggunakan sebuah metode *confusion matrix* dalam proses evakuasi untuk mengetahui hasil akurasi dalam proses klasifikasi. *Confusion matrix* adalah salah satu alat yang penting guna untuk melakukan evaluasi yang digunakan pada *machine learning* yang biasanya memiliki dua kategori maupun lebih (Gumilang, 2018). Setiap unsur matriks menunjukkan jumlah contoh data uji untuk nilai kelas yang sebenarnya yang digambarkan dalam bentuk baris, sedangkan untuk kolom menggambarkan kelas yang diprediksi.

Penelitian ini menggunakan metode *confusion matrix* dalam proses evaluasi mencari nilai akurasi, nilai presisi dan nilai *recall* yang terbaik dalam iterasi dengan menggunakan 2 perbandingan data latih dan data uji, yaitu 80% : 20% dan 70% : 30%.

**Tabel 4. 8 Hasil Confusion matrix**

Perbandingan	Prediksi	Aktual		Akurasi	Presisi	Recall
		Positif	Negatif			
Data uji & data latih 80% : 20%	Positif	1.504	103	83,75%	93,59%	87,13%
	Negatif	222	172			
Data uji & data latih 70% : 30%	Positif	2226	165	83,17%	93,09%	86,74%
	Negatif	340	270			

Pada nilai *confusion matrix* yang ditampilkan pada **Tabel 4.9** terbentuk menggunakan metode *naïve bayes classifier* pada ulasan aplikasi *dana* pada situs *google play* memperoleh hasil dalam bentuk nilai prediksi. Pada perbandingan data latih dan data uji sebesar 80% : 20% terdapat kelas negatif 275 ulasan negatif



aplikasi *dana* terdapat 172 ulasan terklasifikasi dengan benar sedangkan sisanya terdapat 103 ulasan terklasifikasi menjadi kedalam kelas positif. Sedangkan pada kelas ulasan positif aplikasi *dana* terdapat 1.726 ulasan positif aplikasi *dana* terdapat 1.504 ulasan terklasifikasi dengan benar sedangkan sisanya terdapat 222 ulasan terklasifikasi menjadi kedalam kelas negatif dan mendapatkan nilai akurasi tertinggi sebesar 83,75%, nilai presisi sebesar 93,59%, dan nilai Recall sebesar 87,13% .

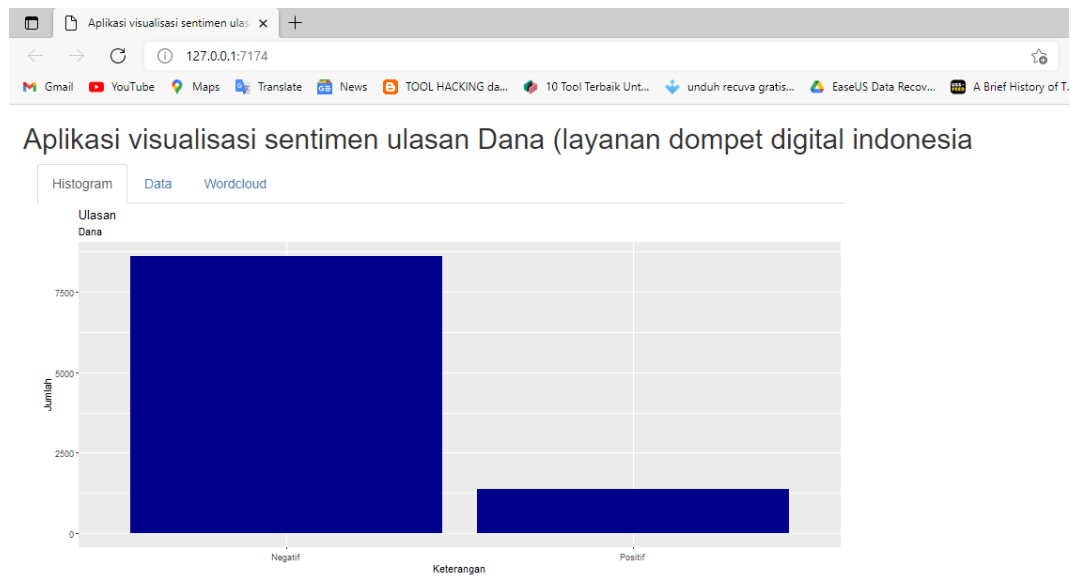
Pada perbandingan data latih dan data uji sebesar 80% : 20% terdapat kelas negatif 435 ulasan negatif aplikasi *dana* terdapat 270 ulasan terklasifikasi dengan benar sedangkan sisanya terdapat 165 ulasan terklasifikasi menjadi kedalam kelas positif. Sedangkan pada kelas ulasan positif aplikasi *dana* terdapat 2.566 ulasan positif aplikasi *dana* terdapat 2.226 ulasan terklasifikasi dengan benar sedangkan sisanya terdapat 340 ulasan terklasifikasi menjadi kedalam kelas negatif dan mendapatkan nilai akurasi tertinggi sebesar 83,17%, nilai presisi sebesar 93,09%, dan nilai Recall sebesar 86,74% .

#### **4.8 Visualisasi dan Asosiasi**

Setelah mendapatkan hasil dari klasifikasi, maka diperlukanya sebuah visualisasi yang bertujuan untuk mengekstraksi informasi secara keseluruhan berupa topik yang paling sering dibicarakan/diulas oleh pengguna aplikasi *dana*, sehingga dari beberapa teks ulasan yang cukup banyak tersebut dapat diambil informasi yang dianggap cukup penting dan dapat digunakan untuk sarana acuan untuk pemilik perusahaan aplikasi *dana* yaitu PT. Espay Debit Indonesia Koe untuk memperbaiki dan menjaga kinerja aplikasi *dana*. Dalam penelitian ini visualisasi hasil dari analisis klasifikasi akan menggunakan *wordcloud*. *Wordcloud* adalah representasi dari suatu data yang menampilkan kumpulan kata-kata penting dan sering muncul dari data tersebut. Semakin besar kata itu muncul dalam *wordcloud*, maka semakin besar pula frekuensi kemunculan kata itu dalam data.

#### 4.8.1 Visualisasi aplikasi

Pada visualisasi aplikasi, menjelaskan mengenai sebuah visualiasasi pada program aplikasi sederhana yang dalam proses pembuatan dan perancangan menggunakan package *shiny* dengan menggunakan bahasa pemrograman R yang hasil outputnya dapat ditampilkan menggunakan website browser *localhost* yang disediakan oleh *Rstudio*.berikut adalah tampilan program aplikasi sederhana tersebut .



**Gambar 4. 15** Tampilan aplikasi

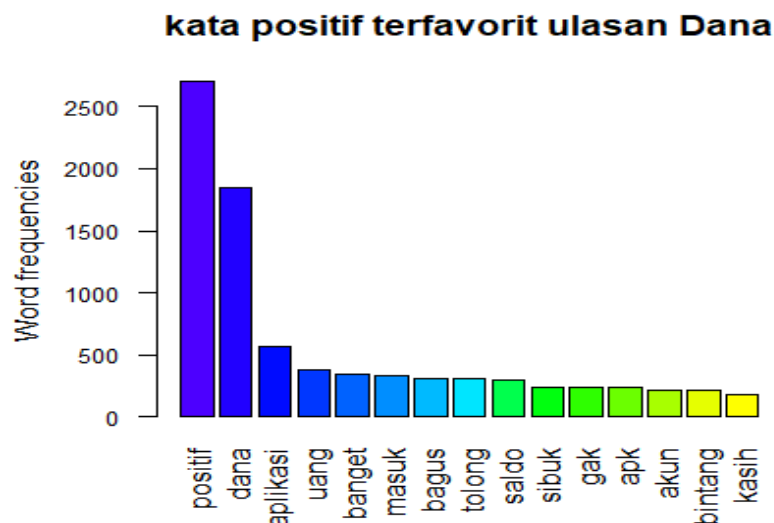
Pada **Gambar 4.15** menampilkan mengenai tampilan sederhana aplikasi visualisasi *Sentiment analyst* yang terdiri dari 3 menu yaitu menu Histogram menampilkan detail jumlah sentiment ulasan yang bersifat positif maupun negatif dalam bentuk diagram batang, selanjutnya terdapat menu data yang berisi hasil data sentimen yang bersifat positif maupun negatif serta jumlah nilai dalam satu kalimat, dan terakhir adalah terdapat menu *wordcloud* yang menampilkan hasil diagram *wordcloud* secara keseluruhan.



Pada **Gambar 4.17** diatas menampilkan mengenai menu *wordcloud* yang menjelaskan tentang keseluruhan kata yang sering muncul dalam ulasan pengguna aplikasi Dana.

#### 4.8.2 Ulasan Positif

Data ulasan positif yang digunakan dalam penelitian ini adalah pelabelan menggunakan analisis sentimen berbasis *lexicon based*. Ekstraksi informasi pada ulasan positif dilakukan secara berulang-ulang hingga mendapatkan informasi tentang ulasan positif mengenai pengguna aplikasi *dana*. Ulasan positif tersebut diidentifikasi berdasarkan frekuensi kata dalam ulasan, berikut adalah visualisasi hasil ekstraksi informasi yang didapatkan dari ulasan pengguna aplikasi *dana* dengan klasifikasi ulasan positif.



**Gambar 4. 18** Kata yang paling banyak muncul pada kelas positif

Pada hasil klasifikasi ulasan positif berbahasa Indonesia , dari jumlah ulasan positif sebanyak 1.383 ulasan, diperoleh beberapa kata yang paling banyak muncul diantaranya adalah kata “positif” dengan frekuensi sebanyak 2.699 kali, “dana” sebanyak 1.841 kali, “bagus” 309 kali, “banget” 343 kali dan seterusnya. Kata-kata yang muncul seperti pada **Gambar 4.18** merupakan kata yang memiliki sentimen positif berbahasa Indonesia dan merupakan topik pembicaraan yang paling banyak diulas oleh pengguna aplikasi *dana* pada situs *google play*. Kata-kata tersebut

selanjutnya digunakan sebagai dasar untuk menemukan asosiasi dengan kata lainnya, sehingga dapat diperoleh informasi yang lebih baik. Kumpulan kata-kata yang sering muncul tersebut dapat ditampilkan dalam bentuk *wordcloud* seperti terlihat pada **Gambar 4.19** berikut.



**Gambar 4. 19** Wordcloud ulasan positif berbahasa Indonesia

Pada visualisasi *wordcloud* dapat dilihat dengan lebih jelas topik dan kata-kata positif apa saja yang sering digunakan pengguna aplikasi *dana* dalam memberikan ulasan. Semakin besar ukuran kata pada *wordcloud* menggambarkan semakin tinggi pula frekuensi kata tersebut, artinya semakin sering pengguna menggunakan kata tersebut sebagai topik pembicaraan atau penilaian positif dalam ulasan. *Wordcloud* pada **Gambar 4.19** merupakan *wordcloud* pada ulasan positif yang selanjutnya, akan dilakukan pencarian asosiasi antar kata yang sering muncul secara bersamaan dan diperoleh hasil sebagai berikut:

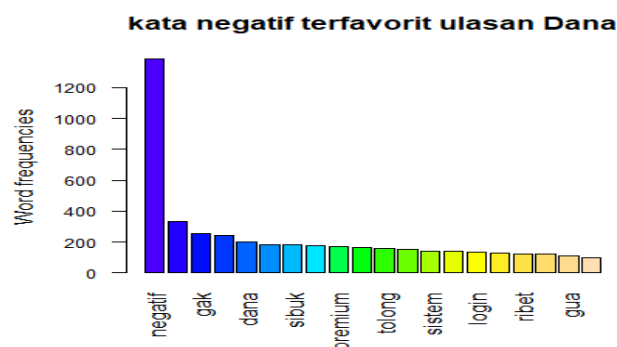
**Tabel 4. 9** Asosiasi pada klasifikasi positif bahasa Indonesia

Dana	Cepat	Jaringan	Bagus
Lengkap (0.26)	Mudah (0.21)	konfirmasi(0.28)	banget (0.23)
Ulasan (0.26)	Akurat (0.18)	kedepan (0.28)	bialng (0.15)
Akun (0.24)	Proses (0.17)	kesalah (0.28)	bisain (0.15)
Saldo (0.22)	Pengirima(0.17)	Barusan (0.28)	percaya (0.15)

Tolong (0.20)	Simple (0.16)	Kalasaham (0.28)	isikan (0.15)
Nonton (0.17)		Padalah (0.28)	knpah (0.15)
Masuk (0.16)		Stabil (0.27)	hadi (0.15)
Mengakses(0.16)		mendaftar (0.23)	iki (0.15)
Pake (0.16)		bermasalah (0.22)	perkenalkan(0.15)
Game (0.15)		kaya (0.22)	pie (0.15)
		tanks (0.20)	
		kenap (0.20)	
		nyah (0.20)	
		sibuk (0.18)	
		login (0.17)	
		koneksi (0.17)	
		pindah (0.17)	

### 4.8.3 Ulasan Negatif

Data ulasan positif yang digunakan dalam penelitian ini adalah pelabelan menggunakan analisis sentimen berbasis *lexicon based*. Ekstraksi informasi pada ulasan negatif ini dilakukan secara berulang-ulang hingga mendapatkan informasi tentang ulasan negatif mengenai pengguna aplikasi *dana*. Ulasan negatif tersebut diidentifikasi berdasarkan frekuensi kata dalam ulasan, berikut adalah visualisasi hasil ekstraksi informasi yang didapatkan dari ulasan pengguna aplikasi *dana* dengan klasifikasi ulasan negatif.



**Gambar 4. 20** Kata yang paling banyak muncul pada kelas negatif

Pada hasil klasifikasi ulasan negatif berbahasa Indonesia , dari jumlah ulasan negatif sebanyak 8.618 ulasan, diperoleh beberapa kata yang paling banyak muncul diantaranya adalah kata “negatif” dengan frekuensi sebanyak 1.384 kali, “sibuk” sebanyak 180 kali, “ribet” 122 kali, “premium” 169 kali dan seterusnya. Kata-kata yang muncul seperti pada **Gambar 4.20** merupakan kata yang memiliki sentimen negatif berbahasa Indonesia dan merupakan topik pembicaraan yang paling banyak di ulas oleh pengguna aplikasi *dana* pada situs *google play*. Kata-kata tersebut selanjutnya digunakan sebagai dasar untuk menemukan asosiasi dengan kata lainnya, sehingga dapat diperoleh informasi yang lebih baik. Kumpulan kata-kata yang sering muncul tersebut dapat ditampilkan dalam bentuk wordcloud seperti terlihat pada **Gambar 4.21**.



**Gambar 4. 21** Wordcloud ulasan negatif berbahasa Indonesia

Pada visualisasi *wordcloud* dapat dilihat dengan lebih jelas topik dan kata-kata negatif apa saja yang sering digunakan pengguna aplikasi *dana* dalam memberikan ulasan. Semakin besar ukuran kata pada *wordcloud* menggambarkan semakin tinggi pula frekuensi kata tersebut, artinya semakin sering pengguna menggunakan kata tersebut sebagai topik pembicaraan atau penilaian negatif dalam ulasan. *Wordcloud* pada **Gambar 4.21** merupakan *wordcloud* pada ulasan negatif yang selanjutnya, akan dilakukan pencarian asosiasi antar kata yang sering muncul secara bersamaan dan diperoleh hasil sebagai berikut:

**Tabel 4. 10** Asosiasi pada klasifikasi negatif bahasa Indonesia

Dana	Aplikasi	Sibuk	Pelayanan	Buruk
Lengkap (0.24)	Bangkrut (0.21)	Sistem (0.73)	Pembatalan(0.30)	Pencairan(0.24)
Ulasan (0.24)	Haram (0.19)	Capek (0.27)	Tutup (0.24)	Bertransaksi (0.17)
Playanan (0.23)	Persulit (0.14)	Review (0.19)	Costumer (0.28)	Pelit (0.17)
Saranin (0.23)	Sejelek (0.20)	Chat (0.18)	Error (0.21)	Promo (0.17)
Sukses (0.20)	Kesusahan(0.14)	Server (0.16)	Lamban (0.15)	Enakan (0.17)
Respon (0.19)	Bosan (0.13)	Kesal (0.15)	Hangus (0.21)	Buriiikkk (0.17)
Curang (0.16)	Burikkkk (0.13)	Panggilan(0.15)	Jelek (0.14)	Pelayananya (0.13)
Jelekkkkkk(0.16)	Beban (0.13)	Login (0.14)	Lambat (0.14)	Lemot (0.10)
Culas (0.11)	Eror (0.12)	Pulsaku (0.14)	Solusi (0.13)	Service (0.11)
Idiot (0.11)	Jelek (0.11)	Daftar (0.13)	Gajelas (0.11)	Melaporkan (0.12)
Begoin (0.11)		Sistem (0.13)	Bintang (0.18)	Terpaksa(0.12)
Keluhan (0.11)		Informasi(0.13)	Turun (0.12)	Pemeliharaan (0.12)
Ribet (0.11)		Nipu (0.11)		Kendala (0.11)
		Telfon (0.12)		
		Bayar (0.10)		

Pada **Tabel 4.10** menjelaskan mengenai kata yang saling berasosiasi pada ulasan kata yang bersifat negatif, Kata-kata tersebut menjelaskan topik yang saling dibicarakan pengguna aplikasi *dana* pada ulasan situs *google play*. Berdasarkan informasi tersebut diperoleh beberapa informasi sebagai berikut.

Kata yang berasosiasi dengan kata “*Dana*” pada ulasan negatif memberikan informasi bahwa respon pelayanan *costumer service* aplikasi *dana* cukup lambat, buruk, dan sulit. Karena ketika pengguna terdapat permasalahan dalam penggunaan aplikasi, pengembalian uang dan transfer, tidak cepat ditangani. Serta permasalahan yang menjelaskan bahwa aplikasi *dana* culas (curang), mungkin dikarenakan respon *costumer service* ketika pengguna mengalami permasalahan dalam melakukan



transaksi menggunakan aplikasi *dana* seperti pembelian maupun transfer, Responya yang kurang sigap dan cepat sehingga menyebabkan pengguna kecewa dan mengira bahwa aplikasi *dana* berbuat curang.

Kata yang berasosiasi dengan kata “*Aplikasi*” pada ulasan negatif memberikan informasi bahwa menjelaskan mengenai keluhan pengguna aplikasi *dana* bahwa aplikasi *dana* yang semakin lama semakin berat yang menyebabkan memerlukan kapasitas penyimpanan yang cukup besar untuk menggunakan aplikasi *dana* sehingga saat digunakan cukup sulit dan sering erorr aplikasinya. Serta permasalahan tampilan aplikasi *dana* yang membosankan sehingga mungkin diperlukanya perbaikan pada sistem dan tampilan pada aplikasi *dana*.

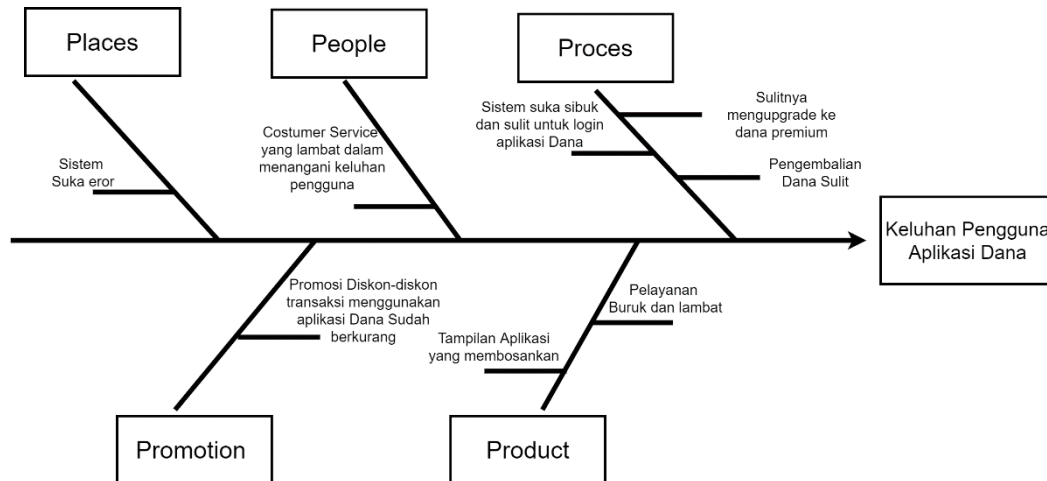
Kata yang berasosiasi dengan kata “*Sibuk*” pada ulasan negatif memberikan informasi bahwa menjelaskan keluhan pengguna aplikasi *dana* mengenai kesulitanya pengguna dalam melaporkan keluhan ke *costumer service* aplikasi *dana* selain responya lambat ketika dihubungi selalu sibuk dan memerlukan pulsa yang cukup besar ketika menghubungi *costumer service* tersebut.

Kata yang berasosiasi dengan kata “*Pelayanan*” pada ulasan negatif memberikan informasi bahwa, menjelaskan keluhan pengguna aplikasi *dana* mengenai Pelayanan ketika pengguna melakukan transaksi menggunakan aplikasi *dana* sering mengalami eror dan pembatalan, ketika pengguna melakukan laporan ke *costumen sevice* responya cukup lamban. Munkin bisa diperbaiki pelayananya dalam menangani keluhan pengguna.

Kata yang berasosiasi dengan kata “*Buruk*” pada ulasan negatif memberikan informasi bahwa, menjelaskan keluhan pengguna aplikasi *dana* mengenai proses pencairan dana (refund) yang lambat suka *error* dan sudah berkurangnya promosi seperti diskon-diskon *merchant* yang biasanya diberikan oleh aplikasi *dana* kepada pengguna. Sarannya mungkin lebih ditambahkan kembali promosi terhadap *merchant* yang sudah berkerja sama dengan aplikasi *dana* sehingga pengguna lebih puas dalam mengguna aplikasi *dana*.

#### 4.9 Diagram Fishbone

Berdasarkan hasil ulasan negatif yang didapatkan dari data ulasan, maka dapat diperoleh sebuah informasi mengenai permasalahan yang terjadi mengenai ulasan negatif aplikasi *dana* pada situs *google play*. **Gambar 4.22** adalah menjelaskan mengenai diagram *fishbone*.



**Gambar 4. 22** Diagram Fishbone keluhan pengguna Dana

Pada **Gambar 4.22** dapat diketahui sebuah informasi factor-faktor yang menyebabkan aplikasi *dana* yaitu dari segi *Process*, *people*, *places*, *product*, *promotion*. Langkah selanjutnya yaitu menentukan pemecahan masalah yang ada pada aplikasi tersebut. Adapun rencana pemecahan masalah pada aplikasi *dana* dapat dilihat pada **Tabel 4.11** Berikut.

**Tabel 4. 11** Rencana pemecahan masalah pada aplikasi Dana

No	Faktor	Permasalahan	Pemecahan Masalah
1	<i>Process</i>	Pengembalian dana uang Sulit (refund)	Pihak <i>Help Center/ Costumer Service</i> segera memperbaiki sistem mengenai fitur pengembalian uang (refund) sehingga menjadi cepat, cekatan dan tidak mempersulit dalam proses pengembalian uang pengguna aplikasi <i>dana</i>

2		Sulitnya <i>mengupgrde</i> ke fitur <i>dana premium</i>	Pihak Dveloper bisa memperbaiki sitem pengguna agar lebih mudah dalam mengupgrade kepada menu fitur <i>premium</i>
3		Sistem suka sibuk dan sulit login pada aplikasi Dana	Pihak Pemilik aplikasi mungkin bisa menambahkan sumber daya manusia yang mempuni agar kita proses aplikasi berjalan tidak mengalami kendala seperti sistem sibuk dan kesulitan dalam <i>login</i> aplikasi <i>dana</i>
4	<i>People</i>	<i>Costumer service</i> lambat dalam menangani keluhan pengguna	Pihak <i>Costumer Service</i> Aplikasi <i>dana</i> mungkin bisa memperbaiki mekanisme dalam proses alur pelayanan keuhan pada aplikasi <i>dana</i> sehingga dalam proses pelayananya semakin cepat dan mudah.
5	<i>Places</i>	Sistem suka <i>eror</i>	Pihak <i>developer</i> bisa memperbaiki sistem aplikasi <i>dana</i> agar ketika pengguna mengguna aplikasi <i>dana</i> tidak mengalami <i>eror</i>
6	<i>Promotion</i>	Promo diskon dalam melakukan transaksi sudah berkurang.	Menambahkan dan memperbanyak kembali program-program diskon dengan merchant yang sudah berkeja sama dengan aplikasi <i>dana</i> sehingga pengguna lebih puas dalam menggunakan dan melakukan transaksi menggunakan aplikasi <i>dana</i>
7	<i>Product</i>	Tampilan Yang <i>Interface</i> aplikasi <i>dana</i> yang membosankan	Memperbaiki tampilan ataupun menambahkan berbagai macam fitur pendukung seperti <i>game</i> , <i>kuis</i> , <i>dll</i> sehingga pengguna aplikasi semakin nyaman dalam mengguna aplikasi tersebut.
8		Pelayanan Buruk dan lambat	Pihak <i>Costumer Service</i> Aplikasi <i>dana</i> mungkin bisa memperbaiki mekanisme dalam proses alur pelayanan keuhan pada aplikasi <i>dana</i> sehingga dalam proses pelayananya semakin cepat dan mudah.

## **BAB V**

### **PENUTUP**

#### **5.1 Kesimpulan**

Berdasarkan hasil analisis dan pembahasan yang telah dijelaskan pada bagian sebelumnya, maka diperoleh beberapa simpulan sebagai berikut

1. Hasil klasifikasi sentimen menggunakan metode *Naïve Bayes Classifier* pada aplikasi Dana dengan menggunakan 8000 data latih dan 2001 data uji. Dari data uji tersebut sebanyak 277 data masuk kelas positif dan sisanya 1.723 data masuk pada kelas negatif atau dengan perbandingan data latih dan data uji sebesar 80% : 20%. serta menghasilkan tingkat akurasi sebesar 83,75%, sedangkan pada saat dilakukan uji coba dengan perbandingan data latih dan data uji sebesar 70% : 30% atau 7000 data latih dengan 3001 data uji. Dari data Dari data uji tersebut 968 data masuk kelas positif dan sisanya 6.033 data masuk pada kelas negatif atau dengan perbandingan data latih dan data uji sebesar. Serta menghasilkan tingkat akurasi sebesar 83,17%.
2. Dengan menggunakan metode asosiasi dan diagram *fishbone* dapat menyelesaikan permasalahan keluhan negatif pengguna, seperti berikut salah satu contoh penyelesaian masalah dalam faktor process seperti pengembalian dana dalam bentuk uang (*refund*) sulit dalam proses pemecahan masalahnya yaitu dengan memperbaiki sistem mengenai fitur pengembalian uang (*refund*) sehingga menjadi cepat, cekatan dan tidak mempersulit dalam proses pengembalian uang pengguna aplikasi dana.
3. Dengan menggunakan metode implementasi visual dengan *wordcloud* peneliti dapat mengetahui topik apa saja yang sering jadi bahan muncul dan menjadi pembicaraan pengguna dalam bentuk kata-kata yang divisualisasikan dengan *wordcloud*.

4. Peneliti dapat menyimpulkan dengan penelitian ini peneliti dapat mengetahui ulasan mana yang bersifat positif, maupun ulasan yang bersifat negatif. Dengan detail ulasan sentimen yang bersifat positif sebesar 1.383 ulasan positif dan ulasan sentimen yang bersifat negatif berjumlah sebesar 8.618 ulasan negatif jika ditotalkan menjadi 10.001 ulasan dengan sentimen positif maupun negatif

## 4.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan maka ada beberapa saran yang disampaikan diantaranya sebagai berikut.

1. Sistem pelabelan kelas sentimen hanya berdasarkan pada susunan kata yang terdapat pada *lexicon based*, sehingga kata-kata negasi belum dapat teridentifikasi dengan baik, untuk penelitian selanjutnya sebaiknya dapat menggunakan sistem pelabelan yang memiliki tingkatan lebih tinggi, sehingga dapat mampu mendeteksi sentimen pada frasa dan kalimat.
2. Untuk penelitian selanjutnya, sebaiknya dapat menggunakan pendekatan machine learning lain sebagai pembandingan performa algoritma *Naïve Bayes Classifier*.
3. Pada penelitian ini data ulasan yang digunakan masih terbatas yang menggunakan bahasa indonesia, untuk penelitian selanjutnya sebaiknya bisa menggunakan data ulasan dengan berbagai macam bahasa.
4. Pada penelitian ini data yang diambil terbatas hanya per-periode 20 April 2021–25 Mei 2021, Untuk penelitian selanjutnya mungkin pengambilan ulasan datanya bisa lebih banyak dan menyeluruh sehingga hasil ekstraksi data yang didapatkan semakin baik.