

## **BAB II**

### **TINJAUAN PUSTAKA & LANDASAN TEORI**

#### **2.1 Tinjauan Pustaka.**

Berikut ini uraian hasil penelitian yang berkaitan dengan penerapan algoritma *Decision Tree & Naive Bayes* terhadap kepuasan pelanggan.

##### **2.1.1 “Implementasi Multinomial Naïve Bayes Untuk Klasifikasi Sentimen Terhadap Pelayanan Perusahaan Otobus Menggunakan Data Facebook”.**

Permasalahan pada penelitian ini dimana dengan seiring berkembang teknologi saat ini upaya dalam mengetahui tingkat kepuasan para pelanggan bisa dilakukan dengan komputerisasi, sehingga tingkat perhitungan kepuasan pelanggan mendapatkan hasil yang lebih akurat dengan hadirnya teknologi komputerisasi saat ini (Jacarria Pangestu & Kodar, 2022). Penelitian ini menggunakan data mining dengan menghubungkan jejaring sosial media untuk mendapatkan dataset, dimana sentimen terdiri dari pelayanan, produk dan perusahaan yang akan dikategorikan menjadi 3 kelas yakni positif, negatif dan neutral yang didapatkan dari salah satu media sosial Facebook. Peneliti mengharapkan dimana tahapan analisis yang baik dan akurat supaya dapat memaparkan informasi yang berharga terkait opini masyarakat, kemudian akan dilakukan pengumpulan menjadi data untuk dilakukan analisis sentimen.

Penelitian ini bertujuan dimana peneliti ingin melakukan analisis terkait informasi yang didapat terhadap perusahaan otobus supaya menjadi acuan dalam mengukur kualitas pelayanan di masa akan datang. Pada penelitian ini menggunakan algoritma *Multinomial Naive Bayes* dengan teks mining data dari media sosial facebook terkait memprediksi sentimen pelayanan terhadap perusahaan otobus.

Hasil dari penelitian ini dimana data diolah menggunakan algoritma *Multinomial Naive Bayes* dapat disimpulkan bahwa kinerja sistem sangat baik dalam menjalankan tahapan klasifikasi pelayanan terhadap perusahaan otobus. Hal ini dapat dilihat dari, hasil persentase data sentimen yang diperoleh dari media

sosial facebook terkait sentimen pelayanan terhadap perusahaan otobus mendapatkan persentase sentimen negatif sebesar 23% , sentimen netral sebesar 53%, dan sentimen Positif sebesar 25% serta pengujian data pada algoritma Multinomial *Naive Bayes* didapat nilai akurasi sebesar 95%, precision sebesar 95%, dan recall sebesar 95%.

Saran untuk penelitian yang akan dilakukan selanjutnya supaya menggunakan media sosial dan menggunakan algoritma lainnya sebagai pembanding agar didapat hasil yang lebih baik dalam tahapan klasifikasi sentimen.

### **2.1.2 “Komparasi Algoritma *Naive Bayes* Dan Algoritma C4.5 Dalam Klasifikasi Pelanggan Produk Indihome”.**

Permasalahan yang ada pada penelitian ini dimana banyaknya persaingan antar perusahaan yang menjalankan bisnis dibidang internet dan langgan jaringan berbayar semakin pesat seiring dengan era perkembangan teknologi saat ini. PT. Telkom Indonesia yang salah satu bisnis produknya yaitu IndiHome ( Indonesia Digital Home ) diharuskan tetap menjaga kualitas dan layanan yang baik supaya dapat mempertahankan pelanggan dan mendapat pelanggan baru yang potensial. Maka dari itu, Indihome memiliki kesulitan dalam mengolah data secara akurat untuk mengklasifikasikan pelanggan potensi dan tidak berpotensi (Syafii et al., 2022).

Penelitian ini memiliki tujuan yakni dimana Indihome menggunakan analisis data CRIPS-DM untuk memperoleh pola informasi terhadap klasifikasi data pelanggan dengan algoritma C4.5 dan *Naive Bayes* untuk memperoleh tingkat akurasi terbaik, supaya memberikan informasi terhadap pelanggan yang memiliki potensi dan tidak berpotensi mengenai produk baru mereka. Kegiatan tersebut guna untuk dijadikan sumber dalam pengambilan keputusan oleh PT.Telkom Indonesia. Data yang digunakan pada penelitian ini yakni dari proses order pelanggan yang melakukan pemasangan produk Indihome, dimana data tersebut akan disimpan menjadi dataset menggunakan csv. Pada proses klasifikasi pelanggan, table yang digunakan berupa tabel transaksi yang didapat dari proses order pelanggan pemasangan Indihome. Adapun atribut pada tabel yakni terdiri dari tanggal order,

nama customer, nd internet, nd voice, nama alpro, tipe transaksi, sto, kecepatan, tipe layananm datel dan kategori.

Hasil dari penelitian ini dimana data diolah menggunakan software RapidMiner, dari kedua algoritma tersebut dapat disimpulkan bahwa kinerja dari algoritma *Naive Bayes* lebih baik digunakan untuk menentukan pelanggan potensial dan tidak potensial pada klasifikasi pelanggan Indihome, dibandingkan dengan C4.5. Hal ini dapat dilihat dari, pengujian data pada algoritma *Naive Bayes* didapat nilai akurasi sebesar 99.81% dan pengujian data pada algoritma C4.5 didapat nilai akurasi 94.63%.

Saran untuk penelitian yang akan dilakukan selanjutnya supaya menambahkan atribut seperti jumlah pembayaran dan tanggal bayar agar mendapatkan hasil dan pola terbaik untuk mengoptimalkan penggunaan metode algoritma lain sebagai komparasi pada algoritma *Naive Bayes* dan C4.5.

### **2.1.3 “Analysis of Passenger Satisfaction on Airlines Using C4.5 and Naive Algorithm”.**

Permasalahan pada penelitian ini adalah dari banyaknya transportasi, pesawat merupakan sarana paling mudah karna melalui jalur udara. Dari kondisi tersebut para perusahaan maskapai penerbangan memiliki potensi yang sangat penting dimana kualitas pelayanan kepuasan penumpang maskapai sering timbul, karena banyaknya penilaian perusahaan bukan dinilai dari sudut pandang pengguna melainkan dari sudut pandang perusahaan (Hendra Kusuma et al., 2022).

Tujuan pada penelitian ini adalah perusahaan diharuskan membuat strategi dimana program pelayanan menjadi orientasi paling penting terhadap pelanggan dengan memperhatikan komponen kualitas pelayanan, agar upaya perusahaan meningkatkan kualitas pelayanan penumpang secara maksimal dan optimal. Data uji yang digunakan pada penelitian ini didapat dari *website* [www.kaggle.com](http://www.kaggle.com) terdiri dari 500 data, 7 atribut dan 1 label dimana jenis data ini termasuk dalam kategori data *supervised learning*.

Hasil dari penelitian ini dimana data diolah menggunakan software RapidMiner, dari kedua algoritma tersebut dapat disimpulkan bahwa kinerja dari

algoritma C4.5 lebih baik dibandingkan dengan *Naive Bayes* hal ini dapat dilihat dari, pengujian data pada algoritma C4.5 didapat nilai akurasi sebesar 81,19% dan pengujian data pada algoritma *Naive Bayes* didapat nilai akurasi 76,24%.

#### **2.1.4 “Studi Komparasi Metode Klasifikasi *K-Nearest Neighbor* dan *Naive Bayes* dalam Mengidentifikasi Kepuasan Pelanggan Terhadap Produk”**

Permasalahan pada penelitian ini dimana upaya dalam memberikan pelayanan yang baik akan membuat pelanggan merasa senang dan puas. Kepuasan berfungsi untuk pembeda kinerja baik yang diharapkan maupun yang dirasakan . Pada penelitian ini untuk memperoleh kriteria yang menunjang diperlukan klasifikasi terhadap kepuasan pelanggan terhadap produk (Febrian et al., 2022).

Penelitian ini bertujuan untuk melakukan studi komparasi terhadap algoritma KNN dan *Naive Bayes* serta mencari tingkat keakuratan yang terbaik dari kedua algoritma tersebut, dimana untuk menghitung nilai akurasi dan kepuasan pelanggan menggunakan data transaksi produk. Data yang diuji diperoleh dari website [www.kaggle.com](http://www.kaggle.com) sebanyak 800 data, dimana data ini dibagi menjadi dua bagian yaitu data latih sebanyak 601 data, dan data uji sebanyak 199. Atribut yang digunakan yakni terdiri dari nama produk, bintang1, bintang2, bintang3, bintang4, bintang5, dan kualitas produk. Pengujian pada penelitian ini menggunakan metode perhitungan Confusion Matrix.

Setelah dilakukan pengujian pada dataset, dimana dari kedua algoritma tersebut dapat disimpulkan bahwa kinerja dari algoritma KNN lebih baik dibandingkan dengan *Naive Bayes*. Hal ini dapat dilihat dari, pengujian data pada algoritma KNN didapat nilai akurasi sebesar 87.84% dan pengujian data pada algoritma *Naive Bayes* didapat nilai akurasi 68.92%.

#### **2.1.5 “Penerapan Algoritma C4.5 Dalam Pengukuran Tingkat Kepuasan Pelanggan Terhadap Layanan Resto Cepat Saji”**

Permasalahan pada penelitian ini dimana banyaknya jumlah restoran cepat saji dengan ide inovasi yang beragam membuat para pelanggan mudah untuk

menjelajahi aneka ragam kuliner (Purwaningsih & Nurelasari, 2022). Oleh karena itu, perusahaan pada industri kuliner memiliki tantangan untuk membuat pelanggan mereka merasa puas dengan cita rasa yang ditawarkan, pelayanan dan produk makanan yang disediakan. Persaingan antara restoran cepat saji diantaranya untuk meningkatkan kepuasan pelanggan dengan meningkatkan kualitas layanan pada restoran cepat saji. Terdapat 5 kategori untuk mengukur kualitas pelayanan pada restoran cepat saji berdasarkan pada nilai kinerja dengan nilai harapan, yakni *tangible*, *reliability*, *responsiveness*, *assurance*, dan *emphaty*. Pelanggan akan memberi penilaian kepuasan menggunakan kuisioner yang diberikan oleh restoran cepat saji.

Penelitian ini bertujuan untuk memperoleh hasil tingkat kepuasan pelanggan pada salah satu restoran cepat saji di Jakarta, untuk mengambil keputusan dalam upaya meningkatkan layanan guna meningkatkan kepuasan terhadap pelanggan restoran cepat saji.

Pada penelitian ini didapatkan hasil dimana dengan menggunakan algoritma C4.5 menghasilkan nilai akurasi sebesar 95,36%. Dengan adanya pengaruh yang diperoleh dari dataset kepuasan pelanggan yang telah diuji memberikan informasi baru kepada restoran cepat saji, bahwa atribut pelayanan staf, porsi, suasana restoran dan harga merupakan acuan dalam menciptakan kepuasan pada pelanggan.

#### **2.1.6 “Analisis Sentimen Terhadap Ulasan Kepuasan Pelanggan Pada Marketplace Tokopedia Di Jejaring Sosial Twitter Menggunakan Algoritma Naive Bayes”**

Permasalahan pada penelitian ini dimana era teknologi yang pesat ini membeli barang secara online pada *marketplace* sangat diminati oleh masyarakat, salah satunya yakni Tokopedia (Nurhakim et al., 2022). Dalam upaya mengetahui tingkat kepuasan para pelanggan terhadap layanan pelanggan pada halaman mereka. Tetapi tidak sedikit pelanggan memberikan ulasan opini secara real time melalui jejaring sosial media, salah satunya yakni Twitter. Dari ulasan opini pelanggan tersebut akan didapatkan sebuah data, jika data tersebut dapat diolah dengan baik

menggunakan analisis sentimen terhadap ulasan opini pelanggan akan dapat digunakan sebagai informasi yang berharga.

Penelitian ini bertujuan sebagai tolak ukur terkait kepuasan para pelanggan dengan dilakukannya penelitian terhadap ulasan opini pelanggan tentang Tokopedia pada jejaring media sosial Twitter. Penelitian ini menggunakan metode klasifikasi data mining dengan algoritma *Naive Bayes* dan memprediksi sentimen terkait ulasan opini masyarakat terhadap tokopedia yang didapat dari jejaring media sosial twitter menjadi 2 kategori, yakni positif dan negatif.

Hasil dari penelitian ini dimana data diolah menggunakan algoritma algoritma *Naive Bayes* dapat disimpulkan bahwa kinerja sistem sangat baik dalam menjalankan tahapan klasifikasi pelayanan terhadap sentimen terkait ulasan opini pelanggan dimana memperoleh lebih banyak kategori positif. Hal ini dapat dilihat dari, pengujian data pada algoritma Multinomial *Naive Bayes* didapat nilai akurasi sebesar 79,02%, precision sebesar 80,30%, recall sebesar 77,94% dan specificity sebesar 80,15%.

### **2.1.7 “Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode *Naive Bayes* dan *Support Vector Machine*”**

Permasalahan pada penelitian ini dimana indonesia merupakan salah satu negara yang terkena dampak dari covid-19 dari sekian banyak nya negara di seluruh dunia, dimana virus ini dalam penyeberan nya sangat cepat (Petiwi et al., 2022). Maka dari itu, banyak negara memberlakukan pembatasan sampai karantina pada wilayah masing- masing, terjadinya pembatasan dengan skala yang cukup ketat ini layanan pengantaran makanan ( *food delivery* ) pada saat masa pandemi menjadi salah satu layanan yang cukup diminati oleh masyarakat dan sehingga berkembang dengan pesat. Melalui riset yang didukung *Criminal Law Student Association* ( CLSA ) menyatakan dimana mayoritas 35 % masyarakat lebih memilih gofood dan 20% memilih aplikasi sejenisnya. Gofood sendiri memiliki strategi dimana mereka memberikan penawaran kepada para pelanggan, sehingga calon pelanggan memiliki tingkat keinginan yang tinggi dalam menggunakan aplikasi gofood. Tidak sedikit para pelanggan memberikan ulasan pro dan kontra pada jejaring sosoal

media secara real time, pada penelitian ini menggunakan salah satu media sosial yakni Twitter. Dari ulasan para pelanggan tersebut akan didapatkan kumpulan data yang akan dilakukan analisis sentimen sebagai tolak ukur peneliti dalam menganalisa terkait ulasan opini masyarakat terhadap pelayanan gofood pada media sosial twitter.

Penelitian ini memiliki tujuan dimana dari kumpulan data ulasan opini masyarakat tersebut akan dilakukan pengelompokkan menjadi 3 kelas kategori yakni, positif, negatif dan neutral serta menggunakan metode klasifikasi data mining dengan algoritma *Naive Bayes* dan *Support Vector Machine*. Peneliti mengharapkan supaya bisa mengetahui apakah para pelanggan puas atau tidak terhadap gofood dan dapat melakukan komparasi terhadap kedua algoritma algoritma *Naive Bayes* dan *Support Vector Machine*. Data yang digunakan pada penelitian ini sebanyak 5000 data tweet yang diperoleh dari tanggal 15 – 25 November 2021 menggunakan keyword “Gofood”.

Hasil dari penelitian ini dimana data diolah menggunakan algoritma *Naive Bayes* dan *Support Vector Machine* dapat disimpulkan bahwa kinerja sistem sangat baik dalam menjalankan tahapan klasifikasi pelayanan terhadap pelayanan gofood. Hal ini dapat dilihat dari, hasil persentase data sentimen yang diperoleh dari media sosial twitter terkait sentimen pelayanan terhadap perusahaan otobus mendapatkan persentase sentimen negatif sebesar 2,0% , sentimen netral sebesar 92,8%, dan sentimen Positif sebesar 5,2% serta pada algoritma *Naive Bayes* didapat nilai akurasi sebesar 74,6% dan 91,5% melalui pengujian antarmuka dan algoritma *Support Vector Machine* didapat nilai akurasi sebesar 83% dan 98,5% melalui pengujian antarmuka. Didapat kesimpulan pada penelitian ini bahwa kinerja pada pengujian data menggunakan algoritma *Support Vector Machine* lebih baik dan akurat dari pada algoritma *Naive Bayes*.

### **2.1.8 “Penerapan Algoritma *Naive Bayes* Untuk Analisis Kepuasan Penggunaan Aplikasi Bank”**

Permasalahan pada penelitian ini dimana era teknologi yang sangat pesat , aplikasi berbasis mobile banyak dibuat pada setiap perusahaan untuk mendukung

berjalannya pelayanan dalam bisnis (Sepri et al., 2020). Salah satunya pada dunia perbankan dimana mengharuskan kepentingan aspek security terhadap data pada saat proses layanan secara online dalam melakukan transaksi yang sedang dilakukan. Tidak sedikit para pengguna memberikan ulasan opini dari media yang menyediakan laman ulasan komentar, dimana salah satu media yakni play store pada mobile yang digunakan pada penelitian ini. Tidak sedikit para pengguna memberikan ulasan opini dari media yang menyediakan laman ulasan komentar, dimana salah satu media yakni play store pada mobile yang digunakan pada penelitian ini. Dari ulasan opini pengguna akan didapatkan kumpulan data yang akan dilakukan analisa sentimen terhadap pelayanan aplikasi mobile bank muamalat, kemudian dari data yang diperoleh akan dikelompokkan menjadi 2 kategori yakni negatif dan positif.

Penelitian ini memiliki tujuan dimana dalam melakukan analisa sentimen terhadap pengguna aplikasi mobile bank muamalat pada media playstore, agar mendapatkan tingkat akurasi yang baik dan tepat menggunakan algoritma *Naive Bayes*. Sehingga peneliti mengharapkan supaya hasil tersebut menjadi tolak ukur untuk penilaian bank muamalat sebagai penentuan kebijakan berbasis komputerisasi.

Hasil dari penelitian ini dapat menyimpulkan dimana dari data sentimen analisis yang diperoleh terhadap penggunaan aplikasi mobile bank muamalat menggunakan algoritma *Naive Bayes* dapat dilihat dari nilai akurasi sebesar 87%. Berdasarkan data training yakni 200 ulasan komentar negatif dan 200 ulasan komentar positif, sedangkan data testing yakni 70 ulasan komentar negatif dan 70 ulasan komentar positif.

### **2.1.9 “Tren Marketplace Berdasarkan Klasifikas Ulasan Pelanggan Menggunakan Perbandingan Kernel *Support Vector Machine*”**

Permasalahan pada penelitian ini dimana dengan perkembangan teknologi yang semakin pesat ini masyarakat dominan lebih menyukai melakukan transaksi pembelian secara online dengan *E-commerce* (Rianti et al., 2021). Namun masih banyak pelaku bisnis dimana mereka masih kesulitan dalam pemilihan marketplace

mana yang akan mereka gunakan terhadap produk yang akan mereka tawarkan kepada pelanggan, dikarenakan para pelaku bisnis kurang memahami tren yang sedang ramai dan disukai para masyarakat. Dari keadaan tersebut tidak sedikit pelanggan memberikan ulasan opini mereka melalui jejaring media social, salah satunya yakni twitter. Maka dari itu, perlu dilakukan analisis oleh para pelaku bisnis terkait tren supaya dapat mengetahui dominan pola pelanggan pada suatu produk dan meningkatkan pada penjualan serta forecasting.

Penelitian ini memiliki tujuan dimana melakukan komparasi terhadap kernel *Support Vector Machine* (SVM) menggunakan metode klasifikasi data mining terkait ulasan opini pelanggan marketplace pada media social twitter. Peneliti mengharapkan dimana dari penelitian ini mampu membuat suatu permodelan yang dapat menganalisis tren pada marketplace, sehingga dapat menjadi tolak ukur terhadap para pelaku bisnis dalam menentukan marketplace yang tepat dan akurat serta menjadi penilaian terhadap marketplace supaya dapat memberikan tingkat kepercayaan kepada pelanggan.

Hasil dari penelitian ini dimana dari proses Analisa tren terhadap marketplace pada twitter memperoleh nilai tertinggi yakni *E-commerce* Bukalapak sebesar 49,71%, Shopee sebesar 38,24% dan Tokopedia sebesar 28,84%. Dapat disimpulkan bahwa kinerja kernel algoritma *Support Vector Machine* (SVM) dapat berjalan dengan baik. Hal ini dapat dilihat dari hasil nilai yang diperoleh sebagai berikut, kernel *polynomial* nilai *accuracy* sebesar 91%, *precision* sebesar 91%, *recall* sebesar 91%, dan *f1 score* sebesar 91%. Pada kernel linear dan RBF nilai *accuracy* sebesar 92%, *precision* sebesar 92%, *recall* sebesar 92%, dan *f1 score* sebesar 91%. Maka permodelan dengan hasil terbaik yakni, kernel sigmoid dengan parameter  $=100$ ,  $=0.01$ , dan  $=1$ . Pada model sigmoid nilai *accuracy* sebesar 92%, *precision* sebesar 92%, *recall* sebesar 92%, dan *f1 score* sebesar 92%.

#### **2.1.10 “Analisis Sentimen Ulasan Pelanggan Batik Ecoprint Menggunakan Naive Bayes Classifier Dan KNN”**

Permasalahan pada penelitian ini dimana budaya batik sudah menjadi ciri khas dari indonesia, salah satunya yakni batik ecoprint dimana memiliki ciri yang

unik sehingga menjadikan batik tersebut memikat banyak pelanggan untuk membeli batik ecoprint yang sedang menjadi trend pada dunia fashion (Saurina et al., 2022). Tidak sedikit online store memberikan penawaran terhadap batik ecoprint untuk para calon pelanggan.

Penelitian ini memiliki tujuan dimana upaya mengetahui ulasan pelanggan terhadap batik ecoprint dan menjadi tolak ukur dalam penilaian dalam kepuasan terhadap pelanggan serta memberi pengetahuan kepada para pelanggan bahwa batik ecoprint merupakan jenis batik yang ramah terhadap lingkungan. Maka dari itu, peneliti akan melakukan proses Analisa sentimen untuk mendapatkan sekumpulan data dari media social twitter yang akan menjadi 2 kategori kelas yakni positif dan negatif menggunakan klasifikasi data mining dengan algoritma *Naive Bayes* dan KNN. Data yang digunakan berupa sentiment data tweet sebanyak 5.736 positif dan sentiment data tweet negative sebanyak 1707, serta data training sebanyak 17.197. Kemudian diperoleh 39 data tweet pelanggan mengetahui bahwa batik ecoprint ramah lingkungan, sedangkan sisanya pelanggan memberikan ulasan terkait kualitas dari batik ecoprint.

Hasil dari penelitian ini dimana dari proses Analisa sentiment terhadap batik ecoprint diperoleh kategori netral, dapat dilihat dari ulasan komentar positif menggunakan keyword “Cinta Batik” sebanyak 583 kata, “senang” sebanyak 74 kata dan “Terima Kasih” sebanyak 31 kata. Kemudian, ulasan komentar negative menggunakan keyword “Tidak Senang” sebanyak 31 kata, “Tidak Puas” sebanyak 17 kata, dan “Batik Buruk” sebanyak 8 kata.

Pada penelitian ini dapat ditarik kesimpulan bahwa kinerja algoritma *Naive Bayes* lebih baik, dapat dilihat dari penjelasan berikut. Algoritma *Naive Bayes* nilai klasifikasi presesi sebesar 85%, *recall* 88% dan *f1-score* 88%. Sedangkan algoritma KNN nilai klasifikasi presesi sebesar 76%, *recall* 76% dan *f1-score* 74%.

## 2.2 Landasan Teori.

Dibawah ini merupakan uraian penjelasan dari teori terkait “Penerapan Klasifikasi data mining Untuk Kepuasan Pelanggan Starbucks Menggunakan Algoritma *Decision tree* & *Naive Bayes*” pada penelitian ini.

## **2.2.1 Big Data**

### **2.2.1.1 Variety**

*Variety* merupakan bagaimana cara pengolahan kerumitan dari beberapa macam data, dimana terdiri dari *structured*, *unstructured* dan *semi structured* (Muttaqin Muttaqin & Rahmadini Darwas, 2022).

### **2.2.1.2 Velocity**

*Velocity* merupakan kecepatan dalam membuat data, mengolah dan menganalisis secara kontinu. Hal tersebut berguna untuk membuat kecepatan lebih tinggi secara real-time dan merupakan kebutuhan dalam memanifestasikan streaming data pada bisnis proses, serta pengambil keputusan.

### **2.2.1.3 Veracity**

*Veracity* merupakan ketepatan data agar data dapat dipercaya, berkualitas sehingga dapat diakses dengan baik (Sunday Ade Sitorus & Hesti Umiyati, 2022).

### **2.2.1.4 Validity**

*Validity* merupakan akurasi dan kebenaran data dimana berguna dalam membuat suatu informasi (PanimalarS et al., 2017).

## **2.2.2 Analisis Sentimen.**

Menurut (Fitriyah et al., 2020) analisis sentimen merupakan tahapan untuk mempelajari, mendapat dan pengolahan suatu data text secara otomatis agar memperoleh sekumpulan informasi berupa sentiment yang terdapat pada ulasan opini. Analisa sentimen bertujuan agar dapat mengetahui opini atau eksistensi opini terhadap suatu informasi pada objek, dengan 2 kategori kelas yakni positif atau negative.

Menurut (Nuraliza et al., 2022) analisis sentimen termasuk salah satu penelitian dari machine learning yang memahami terkait ulasan opini berupa text. Dengan melakukan pendekatan dalam mengelompokkan kategori kelas menjadi positif atau negative. Pengumpulan data ulasan terhadap opini tersebut dimulai dari tahapan scrapping data dan labelling supaya memperoleh 2 kategori data positif dan negative.

Menurut (Resti Wardani et al., 2022) sentimen analisis merupakan salah satu metode dimana penambangan ulasan komentar dan opini terhadap suatu objek pada jejaring social media berbasis *website*. Masyarakat dominan menjadikan media social twitter untuk menjadikan tempat memberikan ulasan opini mereka, baik positif maupun negative.

Dari penjelasan diatas peneliti menarik kesimpulan bahwa analisis sentiment merupakan suatu kegiatan dalam mengumpulkan sekumpulan data yang didapatkan dari berbagai platform jejaring media sosial. Proses kegiatan tersebut bertujuan agar dapat melakukan pengelompokan menjadi 3 kategori kelas yakni, positif, negatif dan netral.

### **2.2.3 Kepuasan Pelanggan.**

Kepuasan pelanggan merupakan salah satu parameter yang dapat diukur dan dilihat pada pelayanan dari suatu perusahaan. eksistensi pelanggan tetap dimana salah satu yang membuat perusahaan semakin berkembang dan tetap mendapat keuntungan (Purwaningsih & Nurelasari, 2022)

Kepuasan pelanggan dapat tercapai jika kualitas pelayanan yang ditawarkan dapat terpenuhi dan fasilitas yang diberikan dapat meningkatkan penilaian pada masyarakat untuk memperoleh hasil yang optimal. Kepuasan pelanggan merupakan tingkatan yang memiliki keunggulan dalam upaya memenuhi harapan para pelanggan, maka dari itu upaya dalam mengetahui tingkat kepuasan para pelanggan dapat dilakukan dengan memberikan kualitas pelayanan dengan optimal pada para pelanggan.

Tidak hanya itu saja upaya dalam menentukan keberhasilan pada suatu perusahaan untuk memberikan kepuasan kepada pelanggan, tetapi keberhasilan pada perusahaan sangat berkaitan dengan sistem kerja sama yang baik pada setiap bagian diperusahaan. Maka dari itu, perusahaan harus membuat suasana kerja yang kondusif supaya tercipta hubungan yang harmonis. Sehingga, diharapkan pada setiap bagian pada perusahaan berperan aktif dalam upaya meningkatkan kualitas layanan.

Kualitas pelayanan yang optimal secara garis besar merupakan perwujudan dalam pelayanan yang benar-benar memberikan kepuasan kepada para pelanggan dalam arti pelanggan benar-benar dilayani secara profesional, menggunakan prosedur aman, lancar, sederhana, ada kepastian biaya dan waktu penyelesaian serta kepastian hukum atas pelayanan diberikan. Pengertian pelanggan menurut UU No. 8 tahun 1999, tentang perlindungan konsumen menyatakan bahwa konsumen merupakan setiap orang pemakai barang dan/atau jasa yang tersedia dalam masyarakat, baik bagi kepentingan diri sendiri, keluarga, orang lain, maupun makhluk hidup lain dan tidak untuk diperdagangkan. Para pelanggan akan merasakan kepuasan dan senang apabila diberikan kualitas pelayanan yang baik. Sementara itu, kepuasan memiliki fungsi terhadap perbedaan baik dari kemampuan kerja yang diharapkan maupun yang dapat dirasakan. Jadi, kepuasan pelanggan mencakup ketidaksamaan terkait pentingnya tingkat layanan dengan hasil yang dapat dirasakan.

Hal itu menjadi suatu hal yang sangat penting agar perusahaan tetap mengupayakan kualitas pelayanan yang optimal terhadap para pelanggan. Supaya tetap dapat mengupayakan kualitas pelayanan yang optimal, perusahaan harus melakukan evaluasi serta pengembangan untuk kedepannya. Dengan perkembangan era teknologi yang semakin pesat ini perkembangan yang sangat dinamis dan persaingan antar perusahaan semakin kompetitif dimana mengharuskan perusahaan memiliki rencana yang matang dan terstruktur. Dalam melakukan semua upaya tersebut diperlukan sebuah data.

Di era perkembangan teknologi yang pesat ini, data merupakan salah satu hal yang cukup berdampak pada proses berjalannya suatu perusahaan. Dari data tersebut dapat dilihat dan diambil gambaran saat dilapangan terkait layanan yang diberikan pada para pelanggan, namun dari data saja belum cukup. Perlu dilakukan proses pada data atau pengolahan data, dimana istilah tersebut biasa dikenal dengan data mining.

#### 2.2.4 Data Mining.

Data mining merupakan ekstraksi informasi atau pola yang penting dari database besar. *Knowledge Discovery in Database (KDD)* merupakan istilah data mining yang sering dikenal. Data mining memiliki beberapa ruang lingkup yakni mulai dari *database system, data warehouse, artificial intelligence, statistic, mechine learning, information tetrieval*, dan komputasi tingkat tinggi. Selain ruang lingkup tersebut, data mining juga mencakup ilmu lain seperti *Neural Network, Pengenalan Pola, Spatial Data Analysis, Image Database, Signal Processing* (Simanjuntak et al., 2022).

Data mining adalah salah satu teknologi yang mampu untuk membantu perusahaan memperoleh informasi penting dari data *warehouse* mereka. Dari beberapa aplikasi data mining yang ada lebih berfokus kepada prediksi, dimana mreka membuat prediksi suatu situasi yang terjadi dimasa sebelumnya dari data yang ada [15].

Data mining merupakan proses yang berguna dalam memahami pola atau informasi penting secara tersembunyi dari data sangat besar sekalipun (Fatmawati & Narti, 2022). Oleh karena itu, biasanya data tersebut sudah diubah dan tersimpan pada database. Berikut merupakan metode – metode yang memiliki fungsi sesuai dengan KDD yang ingin dicari pada data yang akan diolah :

- a. *Association* memiliki fungsi dimana sebagai pencarian relasi terhadap suatu kejadian yang bertambah dengan bersamaan.
- b. *Clustering* mempunyai fungsi dimana pengelompokan data dilakukan terkait dengan menyerupai data lain.
- c. *Prediction* memiliki fungsi dimana pandangan untuk tindakan dalam menggambarkan nilai suatu kejadian dimasa akan datang.
- d. *Estimation* memiliki fungsi persamaan dengan prediksi, akan tetapi estimasi memiliki target kelas terhadap variabel *numeric*
- e. *Classification* memiliki fungsi dimana untuk memetakan data pada suatu kelas yang sudah didefinisikan sebelumnya.

Mengolah data dengan teknik data mining dapat menghasilkan banyak pilihan informasi terhadap perusahaan berupa suatu prediksi, klasifikas dan clustering.

### 2.2.5 Klasifikasi

Klasifikasi merupakan dimana suatu proses untuk menemukan model atau fungsi yang menjelaskan dalam pembedaan kelas data atau konsep dengan tujuan sebagai prediksi kelas terhadap data yang tidak diketahui kelasnya. Model yang diturunkan didasarkan pada analisis dari *training* data ( yaitu objek data yang memiliki label kelas yang diketahui ). Model diturunkan dapat digambarkan dalam berbagai bentuk seperti *if-then* klasifikasi, *decision tree* dan sebagainya.

Cara kerja klasifikasi yakni dengan pengelompokan data sesuai data training dan nilai atribut klasifikasi. Pengelompokan tersebut berguna untuk mengklasifikasikan data baru kepada kelompok yang ada, klasifikasi dapat digambarkan menggunakan pohon keputusan ( *Decision tree* ). Dimana setiap node pada pohon keputusan merupakan pernyataan suatu tes terhadap dataset yang akan diolah, sedangkan setiap cabang dari pohon keputusan merupakan hasil dari tes tersebut. Dari pohon keputusan yang terbentuk dapat diinterpretasikan menjadi sekumpulan aturan pada bentuk *IF condition Then outcome*.

Pada situasi tertentu, pengguna ingin mendapat prediksi dari nilai-nilai data yang tidak tersedia atau hilang ( bukan label atau kelas ). Pada kondisi ini nilai data yang akan diprediksi adalah nilai *numeric* ( angka ). Selain itu, prediksi lebih menekankan pada pengenalan trend dari distribusi dari data yang ada. data mining berguna untuk memprediksi apa yang akan terjadi dimasa yang akan datang.

Klasifikasi adalah salah satu metode yang berguna pada industri bidang data mining . Dimana klasifikasi ialah metode yang berguna untuk menemukan fungsi atau model untuk membedakan kelas tertentu yang memiliki tujuan dalam prediksi kelas label dimana kelas label nya belum diketahui. Dimana klasifikasi ialah metode yang berguna untuk menemukan fungsi atau model untuk membedakan kelas tertentu yang memiliki tujuan dalam prediksi kelas label dimana kelas label nya belum diketahui. Pembuatan model algoritma yang dilakukan pada klasifikasi berdasarkan data latih yang ada, sehingga data yang akan diolah diklasifikasikan berdasarkan model tersebut (Handoko et al., 2022).

Dari pengertian data mining diatas dapat disimpulkan klassifikasi merupakan suatu proses pengelompokkan objek dimana mempunyai beberapa karakteristik tetapi tetap memiliki perbedaan dari satu atau lain cara. Tujuan studi pengambil keputusan yakni untuk memprediksi kondisi berdasarkan hasil yang akan didapat. Berikut proses klasifikasi :

- a. Tahapan pembuatan model, dimana pelatihan dilaksanakan menggunakan kumpulan latihan yang telah diketahui sumber namanya.
- b. Tahapan testing, dimana sering digunakan untuk prediksi label selama fase pengujian untuk memeriksa keakuratan model yang dibuat menggunakan informasi yang disiapkan.

#### **2.2.6 Twitter.**

Twitter merupakan social media yang berguna untuk memberikan para pengguna dalam mengirim dan menerima pesan yang berisikan 140 karakter, akan tetapi pada 07 November 2017 twitter melakukan perluasan sehingga menjadi 280 karakter yang bisa dikenal dengan sebutan tweet dari beberapa individu pada seluruh dunia. Twitter diciptakan oleh Jack Dorsey, Biz Stone, Evan Williams, dan Noah Glass di Walk pada tahun 2006 yang diluncurkan pada Juli 2006 .

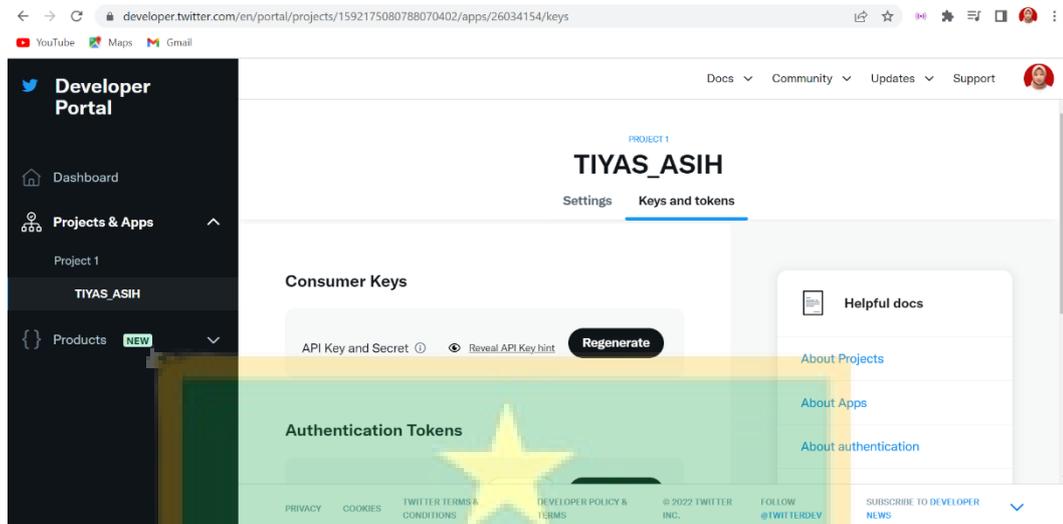


**Gambar 1.** Twitter

Di era perkembangan zaman ini dimana media *social* cukup diminati sebagai penghubung terhadap masyarakat yang salah satunya yakni Twitter (Budianto et al., 2022). Oleh karena itu dapat dilihat dari persentase jumlah peningkatan dalam penggunaan twitter pada seluruh dunia dan indonesia menjadi salah satunya. Penggunaan twitter tidak hanya sebagai media komunikasi dan sosialisasi, tetapi dapat berguna juga sebagai sarana mengutarakan aspirasi serta menggambarkan suatu hal yang sedang terjadi terhadap masyarakat. Dari keadaan tersebut dapat dijadikan sebuah referensi dalam melakukan analisis kepuasan pelanggan terkait Starbucks. Analisis kepuasan pelanggan salah satu cara yang berguna untuk melakukan tahapan proses klasifikasi data mining terkait teks data yang didasari dari opini dan sudut pandang dari kalangan masyarakat yang dilakukan dengan cara memahami, mengolah sampai mendapatkan data yang bersifat data teks secara otomatis. Hal tersebut dilakukan supaya dapat mengetahui apakah opini atau tanggapan dari masyarakat itu positif atau negatif.

### **2.2.7 Twitter Api.**

*Application Programming Integration* (API) twitter adalah suatu aplikasi atau program yang tersedia pada twitter untuk memberikan kemudahan para developer dalam mengakses informasi yang terdapat pada website twitter. Untuk melakukan akses terlebih dahulu para developer melakukan registrasi pada API twitter pada website <https://dev.twitter.com>, sebagai pengguna developer aplikasi twitter. Setelah melakukan registrasi, developer akan memperoleh *consumer key*, *consumer acces*, *acces token* dan *acces token secret* untuk digunakan sebagai syarat otentifikasi pada aplikasi yang akan dibangun. Otentifikasi bertujuan untuk hak akses para *developer* jika ingin mendapatkan data yang terdapat pada twitter (Budianto et al., 2022).



**Gambar 2.** *Developer Twitter Portal.*

Dimana sekumpulan perintah, fungsi dan protokol untuk membuat kode sebagai sistem proses pengambilan image terdiri dari kelas, struktur, fungsi, *interface* sebagai metode pembuatan kode API ( *Application Programming Interface* ). Menurut pendapat salah satu ilmuwan dalam membongkar kode yang kemungkinan dikembangkan menggunakan kode yang berbeda dari API .

### 2.2.8 Confusion Matrix

Penggunaan *Confusion Matrix* merupakan salah satu metode untuk mengukur kinerja pada suatu metode pengelompokkan. Cara kerja *confusion matrix* dilakukan dengan komparasi dataset yang sudah dilakukan pengujian menjadi empat kategori. Istilah yang terdapat untuk menggambarkan hasil dari proses terdiri dari Positif Salah, Negatif Salah, Positif Benar (TP), dan Negatif Benar (TN) .

Dalam dua baris dan kolom *confusion matrix* dirujuk sebagai *true* and *false*.

Keterangan :

TP (*True Positive*) : Jumlah data positif yang diakui dengan benar.

TN (*True Negative*) : Jumlah data negatif yang diakui dengan benar.

FP (*False Positif*) : Jumlah data negatif yang terdeteksi dari data positif.

FN (*False Negative*): Jumlah data positif yang terdeteksi sebagai negatif.

Berikut rumus untuk melakukan pengukuran nilai akurasi untuk klasifikasi text :

$$\text{Recall} = \text{TP}/(\text{TP}+\text{FP}) \quad (1)$$

$$\text{Precision} = \text{TP}/(\text{TP}+\text{FN}) \quad (2)$$

$$\text{Accuracy} = (\text{TP}+\text{TN})/(\text{TP}+\text{TN}+\text{FP}+\text{FN}) \quad (3)$$

### 2.2.9 Algoritma *Decision tree*.

Pada tahapan klasifikasi dataset pada umumnya dapat menggunakan salah satu algoritma untuk memperoleh hasil keputusan yakni Algoritma *Decision tree*. *Decision tree* merupakan salah satu algoritma yang dapat digunakan untuk memprediksi dari data yang sebelumnya sudah tersedia (Tangkelayuk & Mailoa, 2022).

Algoritma *Decision tree* memiliki kelebihan dalam menganalisis logika pada aritmatika seperti pengolahan data *numerik* ( angka ), atribut yang tidak memiliki nilai dan mempunyai aturan yang memperoleh kinerja lebih cepat dibandingkan penggunaan algoritma lainnya .

*Decision tree* atau pohon keputusan adalah salah satu algoritma pada metode klasifikasi data mining yang terkenal dan biasa digunakan pada penelitian. Cara kerja *decision tree*, dengan membuat pohon keputusan yang bertujuan untuk memperoleh hasil dari masalah dari masukan. Algoritma C4.5 dikembangkan untuk membantu dalam pengambilan keputusan membuat serangkaian keputusan melibatkan peristiwa dari ketidakpastian. Algoritma C4.5 dirancang untuk membantu pengambil keputusan membuat berbagai keputusan yang melibatkan ketidakpastian. Algoritma C4.5 adalah salah satu yang dapat digunakan dalam melakukan klasifikasi pohon keputusan atau pembagian yang bersifat prediksi. Klasifikasi pohon keputusan yang terdiri dari node yang berbentuk seperti akar, dan node akar sendiri tidak mempunyai nilai inputan. Node internal merupakan node yang tidak memiliki akar tetapi memiliki satu inputan, sedangkan node lainnya disebut daun. Daun sendiri menggantikan nilai target yang tepat dari salah satu kelas. Berikut langkah-langkah dalam membuat pohon keputusan :

- a. Memilih atribut untuk dijadikan sebagai akar yang berdasarkan dari nilai gain tertinggi dari atribut yang sudah ada. Berikut rumus persamaan untuk menghitung nilai gain :

$$Gain (S,A) = Entropy(s) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy (S_i) \quad (1)$$

Keterangan :

A : Properti

N : Jumlah partisi atribut A

S : Sekumpulan kasus

|S<sub>i</sub>| : Jumlah kasus pada partisi ke-i

|S| : Jumlah kasus dalam S.

Berikut merupakan rumus persamaan untuk menghitung nilai *entropy* :

$$Entropy (s) = \sum_{i=1}^n - p_i * \text{Log}_2 p_i \quad (2)$$

Keterangan:

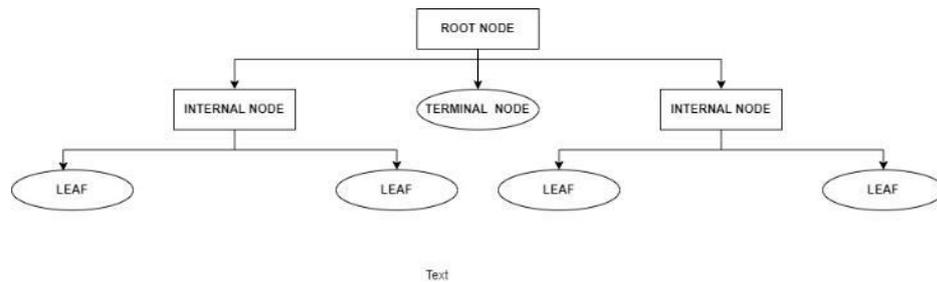
S : Sekumpulan kasus

A : Properti

N : Total Jumlah

p<sub>i</sub> : Rasio dari S<sub>i</sub> terhadap S.

- b. Membuat cabang pada tiap-tiap nilai.
- c. Membagi setiap kasus pada cabang
- d. Membagi kasus pada cabang
- e. Mengulangi setiap tahapan supaya mendapatkan kelas yang sama dari setiap cabang maupun setiap kasus (Fatmawati & Narti, 2022).



**Gambar 3.** Pohon keputusan

Pada pohon keputusan (*decision tree*) diatas terdapat tiga jenis node, yaitu:

1. *Root Node* adalah *node* yang terletak pada bagian paling atas sesudah dilakukannya proses perhitungan nilai *gain* dan *entropy*.
2. *Internal Node* atau *node* percabangan nilai dari setiap atribut, dimana merupakan anak dari struktur *Decision tree*.
3. *Leaf Node* atau *Terminal Node* adalah *node* yang letaknya terdapat dibagian akhir *Decision tree*.

### 2.2.10 Algoritma Naive Bayes.

Algoritma *Naive Bayes* merupakan salah satu algoritma yang sering digunakan pada penelitian dalam upaya mengatasi klasifikasi pada teks dengan menggunakan perhitungan probabilitas [(Djamaludin et al., 2022)].

Algoritma *Naive Bayes* merupakan salah satu algoritma pada klasifikasi data mining yang sering digunakan pada penelitian (Fatmawati & Narti, 2022; Wulandari et al., n.d.). Algoritma *Naive Bayes* dapat melakukan prediksi peluang terjadinya kejadian dimasa akan datang berdasarkan data yang ada, sehingga memperoleh hasil dengan nilai akurasi yang tinggi dan dapat lebih ditingkatkan lagi menggunakan seleksi fitur ( *feature selection* ). Algoritma *Naive Bayes* menggunakan model fitur independen yakni fitur dalam sebuah data yang tidak berkaitan dengan ada atau tidak fitur lainnya pada data yang sama. Keuntungan algoritma *Naive Bayes* dimana algoritma ini hanya memerlukan data latih dengan jumlah kecil untuk memperkirakan nilai parameter (rata-rata dan varian dari variabel) yang diperlukan dalam klasifikasi data mining .

Algoritma *Naive Bayes* adalah salah satu algoritma pada metode klasifikasi data mining dimana bayesian sederhana menggunakan *teorema Bayes* dengan cara menghitung *probabilitas*, algoritma *Naive Bayes* sudah sering digunakan pada penelitian. *Naive Bayes* mempunyai kecepatan tinggi dalam pengaplikasian dengan data yang jumlahnya besar. Berikut adalah rumus persamaan perhitungan algoritma *Naive Bayes* :

$$P(H|Y) = \frac{P(H) \times \prod P(Y_i|H)}{P(Y)} \quad (1)$$

Keterangan:

H : Hipotesis Y untuk kelas tertentu

Y : data kelas yang tidak diketahui

$P(H|Y)$  : Probabilitas hipotesis H (probabilitas posterior) berdasarkan kondisi y

$P(H)$  : probabilitas hipotesis H (probabilitas sebelumnya)

$P(Y_i|H)$  : Probabilitas hipotesis Y mengingat kondisi hipotesis

$P(Y)$  : Probabilitas Y

Keterangan :

Probabilitas posterior : probabilitas bahwa kelas H terjadi

Probabilitas Sebelumnya : Probabilitas sampel awal untuk kelas Y.

Langkah-langkah klasifikasi dengan NBC diuraikan sebagai berikut.

- Pembagian data menjadi 2 macam yakni, data training dan data testing.
- Pada data training dilakukan perhitungan probabilitas prior ( $P(C_k)$ ) dan likelihood( $P(C_k)$ ) setiap variabel.
- Menghitung probabilitas posterior( $P(X_1, \dots, X_n)$ ) pada data testing dimana penentuan probabilitas prior dan likelihood didasarkan pada hasil perhitungan dari data training.
- Memilih kelas dengan probabilitas posterior yang besar untuk menentukan dikelas mana suatu data diklasifikasikan.

### 2.2.11 Python.

Python merupakan salah satu bahasa pemrograman yang dibuat oleh Guido van Rossum (Susanti & Walid, 2022a). Python adalah bahasa pemrograman yang mudah dipahami, karena menggunakan *syntax* yang mudah untuk dimengerti bagi orang yang ingin mempelajari lebih dalam bahasa pemrograman ini. Berikut aspek pemrograman seperti prosedural, *object-oriented programming* dan juga fungsional yang merupakan Bahasa pemrograman yang didukung python. Python juga memiliki banyak pustaka pendukung untuk proyek pembelajaran mesin, termasuk SKLearn, Numpy, dan Pandas.



Gambar 4. [Python](#)

Python merupakan salah satu bahasa pemrograman menggunakan script yang artinya dapat dijalankan tanpa perlu *compiler*. Oleh karena itu banyak developer software dalam membuat program yang sederhana dengan cepat atau untuk membuat tugas yang perlu di eksekusi berlanjut secara otomatis.

Sebelumnya python digunakan oleh para developer dikarenakan kemampuan dalam manipulasi tipe data string (teks). Dengan seiring berkembangnya waktu python dapat digunakan dalam pengolahan data sebagai analisis.

Bahasa pemrograman python cukup banyak dipergunakan dalam bidang machine learning dan data science. Berikut terdapat beberapa manfaat dari python:

- 1) Untuk cleaning data.
- 2) Untuk membuat gambaran ( *visualisasi* )
- 3) Untuk membuat pengembangan suatu model
- 4) Memiliki banyak *library* lengkap
- 5) Bersifat *open source*

- 6) Bersifat *fleksibel*, jadi dapat meningkatkan tingkat pengembangan para *developer*.

### 2.2.12 Anaconda Navigator & Jupyter Notebook.

Anaconda *navigator* merupakan tampilan *Graphical User Interface* ( GUI ) pada dekstop dimana salah satu dari output anaconda sebagai upaya pengguna dalam membuat aplikasi dan mengelola paket, lingkungan dan saluran Conda tanpa menggunakan perintah pada baris(Susanti & Walid, 2022b). Sedangkan, jupyter sendiri berasal dari 3 singkatan bahasa pemrograman kata yakni, Julia(Ju), Python(Py), dan R. definisi dari jupyter notebook ialah salah satu aplikasi free source code yang banyak digunakan oleh para data scientist. Jupyter notebook digunakan dalam pembuatan dan memberikan dokumen yang mempunyai kode, hasil perhitungan, tampilan visualisasi dan teks. Pada penelitian ini peneliti menggunakan jupyter notebook dalam membuat baris perintah dengan bahasa pemrograman python untuk melakukan analisa data dan pengujian data menggunakan metode klasifikasi data mining dengan algoritma *Decission Tree & Naive Bayes*.

Jupyter notebook merupakan lingkungan yang cukup interaktif dalam menjalankan baris perintah berupa kode pada browser dan termasuk tools yang berguna dalam membuat sebuah pekerjaan yang berhubungan dengan analisis ilmuwan data (Jurnal et al., 2022). Jupyter notebook berguna dalam memudahkan inputan berupa kode, teks dan gambar. Kemudian, jupyter notebook memuat bahasa pemrograman tidak hanya python.