

BAB II

TINJAUAN PUSTAKA

2.1 Teori Dasar

Teori dasar adalah sistem yang digunakan untuk mengidentifikasi bagaimana sebuah konsep berhubungan satu sama lain. Teori dasar juga mencakup hasil penelitian dan digunakan oleh peneliti sebagai kerangka teori untuk menyelesaikan penelitian mereka. KDD (*Knowledge Discovery in Database*), data mining, metode Naive Bayes, *software* pendukung, objek penelitian, dan kerangka pemikiran adalah teori dasar dari penelitian ini.

2.2 Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah proses umum yang menggabungkan beberapa metode untuk mengekstrak pengetahuan atau informasi dari basis data yang besar atau kompleks. KDD mencakup semua aspek kehidupan kerja, mulai dari pengumpulan dan interpretasi data hingga interpretasi analisis hasil. Menurut (Handoko & Lesmana, 2018) *Knowledge Discovery in Database* (KDD) mencakup pengumpulan dan penggunaan data historis untuk menemukan keteraturan dan pola hubungan dalam set data berukuran besar. Menurut (Chairil Adam, 2023) KDD adalah metode pengumpulan dan pemrosesan data yang bertujuan untuk mengekstraksi informasi penting yang sudah ada, yang dapat dilakukan oleh perangkat lunak yang menggunakan perhitungan statistik matematis atau kecerdasan buatan. Menurut (Klasifikasi et al., 2023b) KDD adalah jenis ilmu pengetahuan yang mencakup database dan dapat diuraikan sebagai proses pengenalan pola pengetahuan yang menghasilkan kumpulan data yang besar dan mudah dipahami, Ada beberapa tahapan – tahapan pada proses KDD yaitu Data

Cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Pattern Evaluation, Knowledge Presentation. Beberapa proses pengumpulan data KDD termasuk pemurnian data, integrasi data, pemilihan data, dan presentasi pengetahuan. Banyak industri, termasuk bisnis, pendidikan, dan pemerintahan, telah menggunakan pemrosesan data komputerisasi. Komputer yang memiliki data disimpan dalam suatu database data dapat melakukan ini.

2.3 Data Mining

Menurut (Jangra & Kalsi, 2019) data mining adalah proses mengumpulkan semua informasi yang diperlukan dan menggunakan metode statistik untuk analisis data. Menurut (Mandala & Putri, 2023) Data mining adalah salah satu metode yang sangat sering digunakan didalam penelitian, dimana data mining dapat membantu menyelesaikan masalah tentang analisis dan dapat membantu menyelesaikan masalah bisnis, kemampuan data mining dalam menganalisis data dalam jumlah yang besar sudah terbukti didalam berbagai bidang. Berbagai alat bantu augmentasi data telah dikembangkan untuk augmentasi data yang digunakan untuk menganalisis berbagai jenis data. Sistem basis data, pergudangan data, statistika, pembelajaran mesin, pengambilan data, dan komputasi tingkat tinggi adalah bidang keilmuan lain yang berhubungan dengan data mining. (Susana & Suarna, 2022). Data mining yang juga disebut sebagai penemuan pengetahuan dalam database (KDD), adalah proses untuk mengumpulkan informasi berharga dari gudang basis data yang berupa ilmu pengetahuan (Ulfa & Amin, 2020). Selain itu, Data mining adalah kumpulan proses yang dilakukan untuk mendapatkan nilai tambahan dari kumpulan data dengan menggunakan pengetahuan yang sebelumnya tidak dapat diperoleh melalui penggalian manual. (Aranski et al., 2019).

Selain itu, data mining juga didukung oleh ilmu-ilmu lain seperti jaringan syaraf tiruan, pengenalan pola, analisis data spasial, database gambar, dan sinyal. Klasifikasi adalah proses berguna untuk menemukan pola tertentu dari data yang ada di database. Metode ini biasanya digunakan dalam bidang pengajaran mesin dan statistika, di mana informasi diperdalam sehingga membentuk pola-pola yang menarik dari sumber data yang besar yang dapat diperoleh. Dalam praktek pengambilan keputusan, matematika diperlukan untuk data mining untuk mendukung tahap perencanaan, memberikan informasi yang tepat untuk membuat prediksi berdasarkan pengalaman masa lalu untuk kondisi (Klasifikasi et al., 2023). Pada data mining terdapat tiga bagian yaitu asosiasi, klasifikasi, dan clustering. Asosiasi adalah yang mengacu pada definisi proses yang digunakan untuk menemukan hubungan yang ada pada nilai atribut kumpulan data yang memilikinya, proses mencari sekumpulan model yang menggambarkan dan membedakan kelas-kelas data dikenal sebagai klasifikasi. Tujuan dari klasifikasi adalah agar model tersebut dapat digunakan untuk memprediksi kelas objek yang kelasnya belum diketahui (Susana & Suarna, 2022). Teknik yang digunakan untuk menggunakan instance adalah clustering yang mempunyai arti mengelompokkan data ke dalam kelas-kelas objek tanpa mempunyai dasar kelas data tertentu, yang sesuai dengan topik yang sedang dibahas.

2.4 Metode Naive Bayes

Naive Bayes adalah salah satu metode klasifikasi dalam bidang pembelajaran mesin (machine learning) dan teori probabilitas. Metode ini didasarkan pada teorema Bayes, yang menyediakan kerangka kerja untuk menghitung probabilitas suatu kejadian berdasarkan informasi sebelumnya terkait kejadian tersebut. Pada

metode Naïve Bayes terdapat dua kelas yaitu layak dan tidak layak (Klasifikasi et al., 2023). Naïve Bayes merupakan salah satu teknik pembelajaran pada mesin yang perhitungannya menggunakan probabilitas. Konsep dasar yang digunakan dalam Naïve Bayes adalah teorema Bayes yang melakukan klasifikasi dengan menghitung nilai probabilitas p (Susana & Suarna, 2022). Metode Naive Bayes, atau teorema keputusan Bayes adalah pendalaman statistik yang akurat untuk persepsi bentuk. Kuantifikas mendukung keyakinan ini. (Situmorang & Sirait, 2020). Untuk pemodelan prediktif, Naïve Bayes adalah teknik berbasis statistik yang sederhana namun efektif yang bergantung pada teorema Bayesian tentang kemungkinan yang membuat probabilitas untuk setiap peristiwa. Naïve Bayes menganggap bahwa setiap fitur independen, dan karenanya output probabilitas yang paling tinggi diprediksi (Elankath & Ramamirtham, 2023). Klasifikasi Naïve Bayes berfokus pada proses pembelajaran daripada estimasi probabilitas. Selain waktu pengerjaannya yang singkat, strukturnya yang relatif sederhana membuatnya mudah diperkenalkan dan sangat efektif. Naïve Bayes adalah metode klasifikasi text mining yang juga dapat digunakan untuk analisis sentimen dan hanya membutuhkan data pelatihan yang sedikit. Pada Naïve Bayes mengasumsikan bahwa nilai yang di input itu tergantung dengan nilai atribut lain, probabilitas atau disebut peluang bersyarat dinyatakan dengan sebagai berikut. Persamaan umum yang dimiliki pada algoritma klasifikasi Naïve Bayes menurut (Oktavia Putri Zusrotun1, 2022) dapat dilihat pada Persamaan dibawah ini.

Teorema umum pada probabilitas Naïve Bayes:

$$P(H|X) = \frac{P(X|H) * P(H)}{P(X)}$$

Rumus 2.1 Teorema Naïve Bayes

$P(H|X)$ = Probabilitas hipotesis H berdasar kondisi X

$P(X|H)$ = Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$ = Probabilitas kejadian X

$P(H)$ = Probabilitas kejadian H

X = data dengan kelas yang belum diketahui; H = Hipotesis data X merupakan kelas spesifik.

Nilai polaritas data, akurasi klasifikasi, dan matriks sentimen akan dibuat setelah data dilakukan *preprocessing*. Nilai kinerja akan dihasilkan selama proses klasifikasi. Nilai *precision*, *recall*, dan lainnya adalah beberapa alat yang digunakan untuk menghitung nilai manfaat.

$$\text{Precision} = \frac{SY}{SY+ST}$$

$$\frac{JY}{JY+JT}$$

Keterangan:

SY = SeringYa

ST = SeringTidak

JY = JarangYa

JT = JarangTidak

Sedangkan perhitungan nilai Recall untuk class yang beranggapan Jarang dapat dilihat pada persamaan dibawah:

$$\text{Recall} = \frac{SY}{SY+JT+JY}$$

2.5 Software Pendukung

Software pendukung adalah suatu perangkat lunak yang digunakan peneliti untuk mendukung penelitiannya, Dibawah ini merupakan *software* pendukung yang digunakan peneliti:

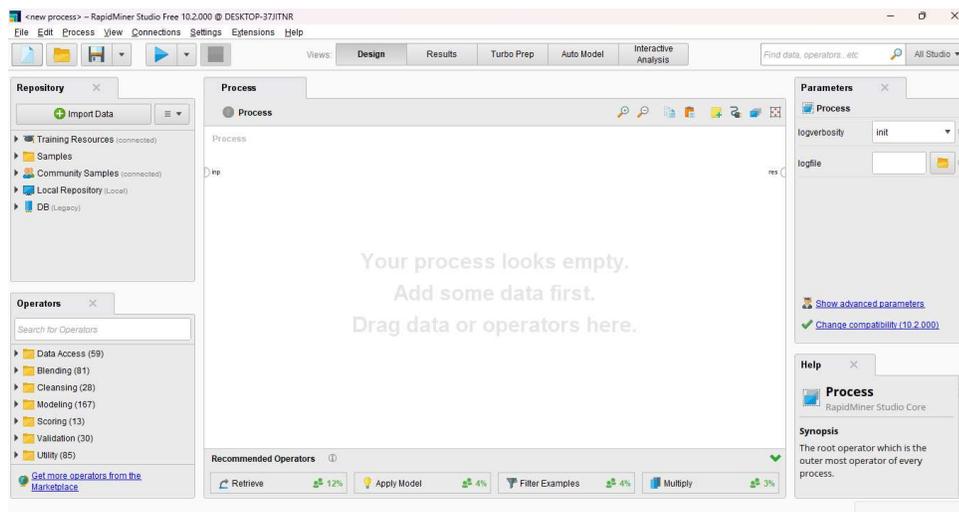
1. RapidMiner

RapidMiner merupakan Metode yang biasanya digunakan untuk mencari solusi data mining karena dapat bekerja di berbagai sistem operasi. Sebagai perangkat lunak sumber terbuka, telah mencapai tonggak sejarah pertamanya sebagai perangkat lunak penambangan data ketika digunakan oleh KDnuggets, sebuah platform penambangan data, pada tahun 2010-2011. RapidMiner yang dikenal sebagai alat untuk membuat keputusan yang tepat, menilai perilaku pengguna dengan teknik prediktif dan deskriptif. RapidMiner adalah perangkat lunak sumber terbuka. Sebelum dikenal dengan nama Rapid Minner, Adapun nama lain dari perangkat ini yaitu YALE (*Yet Another Learning Environment*) yang dikembangkan oleh Ralph Klinkenberg pada tahun 2001.

Kdnuggets mengklaim bahwa RapidMiner adalah program penambangan data terbaik. Untuk desain pipeline yang lebih rinci, RapidMiner menawarkan GUI (*Graphic User Interface*) yang membuat file XML (*eXtensible Markup Language*), yang menentukan metode pengurutan yang harus digunakan pengguna untuk data. Analisa langsung mengaktifkan file setelah RapidMiner membacanya. Berikut beberapa sifat pada RapidMiner antara lain:

1. Java ditulis dengan menggunakan bahasa pemograman sehingga berjalan sesuai sistem operasinya.

2. Model pohon menggambarkan metode pengumpulan data. Struktur internal XML.
3. Eksperimen skala besar dapat dilakukan dengan menggunakan bahasa *scripting*.
4. Perencanaan *multi-level* yang ditingkatkan untuk mempertahankan format data yang efisien dan mengelola penanganan data.
5. Sistem ini mencakup GUI, mode baris perintah, dan Java API yang dapat diunduh untuk memanggil program lain.



Gambar 2.1 Aplikasi RapidMiner Studio

Sumber: RapidMiner Studio (10.2)

Selain itu, RapidMiner juga memiliki beberapa fitur antara lain sebagai berikut:

1. Ada banyak algoritma data mining, seperti pohon keputusan dan peta *self-organisasi*.
2. Ada berbagai bentuk grafis, seperti tumpeng tindih diagram histogram, diagram pohon, dan plot tumpeng tindih 3D.
3. Ada berbagai plugin, seperti plugin teks untuk menganalisis teks.

4. Menyediakan prosedur data mining dan pembelajaran mesin termasuk ETL (*extraction, transformation, load*), pra-pemrosesan data, visualisasi data, pemodelan, dan evaluasi model.
5. Menggabungkan proyek mining dan pembelajaran mesin

2.6 Objek Penelitian

Peneliti akan meneliti dan menganalisis sasaran ilmiah untuk mendapatkan data yang relevan untuk tujuan dan manfaat tertentu. Berikut adalah beberapa objek penelitian yang akan diteliti dan dianalisis oleh peneliti:

2.6.1 Sentimen

Analisis sentimen adalah proses mengidentifikasi sentimen dan mengelompokkan polaritas teks dalam dokumen atau kalimat untuk menentukan kategori sentimen positif, negatif, atau netral (Prabowo & Wiguna, 2021). Analisis sentimen dapat menunjukkan emosi sedih, Bahagia, atau marah (Salsabila et al., 2022). Dalam penelitian, penggunaan Analisis sentimen sendiri juga menggunakan teknik yaitu untuk menyatakan suka atau tidak suka terhadap suatu hal berdasarkan sentimen positif dan negatif. Ini dilakukan sebagai parameter dalam proses pengambilan keputusan (Husada & Paramita, 2021).

2.6.2 Gadget

Gadget adalah alat komunikasi yang serba guna. Alat komunikasi ini merupakan perangkat serbaguna dengan banyak kemampuan. Perangkat ini dinilai lebih lengkap dibandingkan bentuk komunikasi elektronik lainnya karena fungsi dan atributnya yang beragam. Saat ini, banyak *gadget* yang tersebar di seluruh dunia dan perkembangannya sangat luas karena dapat mengakses informasi berbeda

yang diperlukan. Adanya perangkat membuat masyarakat lebih mudah melakukan berbagai tugas yang dulunya sangat sulit dilakukan (Fitriana et al., 2021).

2.6.3 Sekolah

Pada penelitian ini, objek yang akan di jadikan sebagai tempat melakukan penelitian adalah Sekolah SMP NEGERI 27 BATAM di Jl. Kav. Bukit Seroja, Sei Pelunggut, Kec. Sagulung, Kota Batam. sekolah ini didirikan pada tanggal 1 April 2007. SK pendiriannya adalah KPTS.178/HK/IV/2008. Sekolah ini merupakan sekolah yang menerapkan kurikulum belajar pemerintah 2013. SMP Negeri 27 Batam merupakan salah satu sekolah menengah pertama yang beragreditas A dan tersertifikasi BAN-S/M (Badan Akreditasi Nasional) Sekolah/Madrasah dengan nilai 95 (Sertifikasi 2021). SMP Negeri 27 Batam mempunyai 30 ruangan kelas, 1 perpustakaan, 1 ruang laboratorium IPA, dan 1 Ruang lab computer.

2.7 Penelitian Terdahulu

Penelitian ini akan didasarkan pada penelitian sebelumnya dan akan digunakan sebagai bahan referensi untuk menemukan kesamaan dan gagasan baru. Berikut adalah beberapa penelitian sebelumnya yang masih terkait dengan subjek penelitian ini.

1. Menurut penelitian (Bansal et al., 2019) yang berjudul **“Using Data Mining Techniques to Analyze the Customers Reaction towards Social Media Advertisements”** The primary goal of this study is to evaluate consumer sentiment as it relates to Facebook product evaluations and comments. Sentiment analysis is used to categorize comments as neutral, positive, or negative before assigning a number between 0 and 1. To undertake a comparison analysis and determine which classification algorithm has the highest accuracy for predicting online product sales,

this research examines a variety of algorithms, including K Nearest Neighbors (KNN), Support Vector Machine (SVM), and Naive Bayes Classifier. Opinion mining and sentiment analysis are important tools for figuring out how customers feel about items and for informing future developments based on their input, according to research. The words that are present in a phrase are regarded as 1 in the matrix when using the Bag of Word model to extract features from input comments.

2. Pada penelitian yang dilakukan oleh (Jangra & Kalsi, 2019) yang berjudul **“Naive Bayes Approach for the Crime Prediction in Data Mining”** The study examined the KNN and Naive Bayes classifiers' performance in crime prediction analysis with an emphasis on crime prediction in India. The conclusion is that because there are so many attributes in the data set, crime prediction is a complicated challenge. When it comes to crime prediction, the Naive Bayes classifier method is more accurate than the KNN classifier method currently in use. Improved outcomes can be obtained by analyzing the relationship between dependent and independent factors using data mining techniques like Naive Bayes.
3. Pada penelitian yang dilakukan oleh (Zusrotun et al., 2022) ISSN : 2089-8673 (print) | ISSN 2548-4265 (Online) dengan judul **“Sentimen Analisis Belajar Online Di Twitter Menggunakan Naïve Bayes”**. Dalam penelitian ini, menyatakan bahwa Berdasarkan data, rating, dan evaluasi, metode Naïve Bayes Classifier dalam Sistem mengklasifikasikan data menjadi sentimen positif dan negatif. dan komentar tentang pembelajaran online yang dikumpulkan oleh *Twitter*. Klasifikasi yang digunakan menggunakan dua kelas: positif dan negatif. Oleh

karena itu, diharapkan metode Naive Bayes dapat digunakan untuk membangun desain prototipe untuk data analitik. Algoritma Naive Bayes digunakan untuk mengklasifikasikan berdasarkan komentar di Twitter, dan dengan mengklasifikasikan komentar pada penelitian ini kita dapat memahami sejauh mana penyebaran dan pengaruh pembelajaran online melalui media sosial.

4. Pada penelitian yang dilakukan oleh (Situmorang & Sirait, 2020) dengan ISSN (Online) 27156265 97 dengan judul **“Implementasi Data Mining Untuk Memprediksi Kelulusan Siswa Dengan Metode Naïve Bayes”**. Dalam penelitian ini menyatakan bahwa Metode Naive Bayes adalah algoritma klasifikasi yang digunakan untuk menganalisis pola dan membuat prediksi berdasarkan perhitungan probabilistic. Penelitian ini berfokus pada penerapan teknik penambangan data menggunakan metode Naive Bayes. Pengumpulan data diambil dari SMK Putra Jaya School di Batam dan menggunakan aplikasi WEKA untuk menghitung serta mendapatkan hasil kelulusan siswa.

5. Pada penelitian yang dilakukan oleh (Kautsar et al., 2023) ISSN: 2809-9834, DOI: 10.31284/p.semtik.2023-1.3938 dengan judul **“Analisis Sentimen Tanggapan Masyarakat Terhadap Pembelajaran Daring Di Masa Pandemi COVID-19 Melalui Sosial Media Twitter Menggunakan Klasifikasi Naive Bayes”**. Dalam penelitian ini menarik kesimpulan bahwa dengan menggunakan pendekatan Naive Bayes untuk analisis sentimen untuk menganalisis sentimen tanggapan publik terhadap pembelajaran online selama pandemi COVID-19. Algoritma Naive Bayes digunakan untuk mengklasifikasikan sentimen tweet yang dikumpulkan dari situs media sosial seperti Twitter. Penelitian ini bertujuan untuk

mengevaluasi kepuasan masyarakat dengan proses pembelajaran online selama pandemi dengan menganalisis sentimen yang diungkapkan dalam tweet. Keakuratan analisis sentimen diukur, dan penelitian melaporkan tingkat akurasi 80,67%.

6. Pada penelitian yang dilakukan oleh (Verawati & Audit, 2022) ISSN 2614-5278 (media cetak), ISSN 2548-8368 (media online) dengan judul **“Algoritma Naïve Bayes Classifier Untuk Analisis Sentiment Pengguna Twitter Terhadap Provider By.u”** menarik kesimpulan bahwa Analisis sentimen merupakan salah satu teknik Untuk mendapatkan sentiment data yang tidak terstruktur kemudian diproses menjadi data terstruktur. Penelitian ini mengevaluasi kinerja sistem analisis sentimen menggunakan berbagai metrik seperti akurasi, presisi, mengingat, dan F-score. Evaluasi melibatkan pembuatan matriks kebingungan untuk mewakili hasil klasifikasi dari kumpulan data. Penelitian menggunakan metode crawling dengan bantuan Twitter API untuk mengumpulkan data, dan data disimpan dalam format.xlsx.

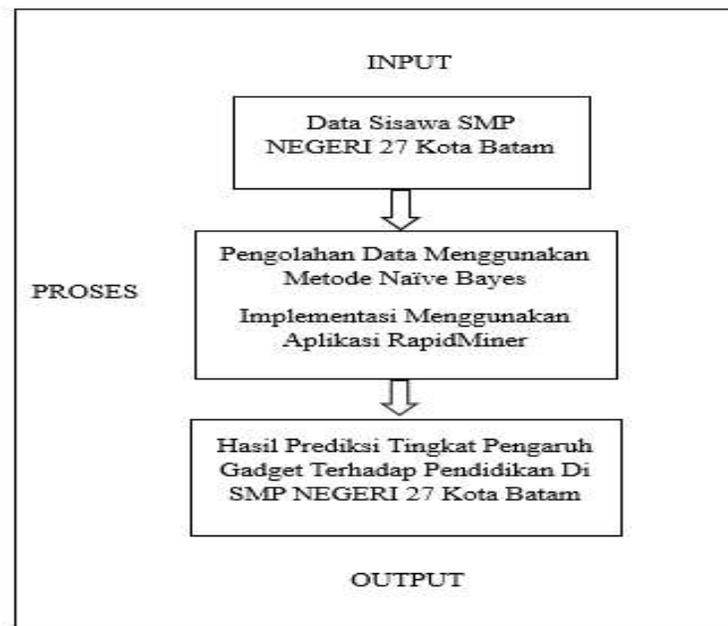
7. Pada penelitian yang dilakukan oleh (Putra et al., 2020) E-ISSN : 2528-4053 | P-ISSN: 2528-4061 dengan judul **“Analisis Sentimen Dompot Elektronik Pada Media SosialTwitter Menggunakan Naïve Bayes Classifier”** Dalam penelitian ini menyatakan bahwa Penelitian ini menggunakan teknik Naive Bayes dan perangkat lunak rapid minner untuk mengevaluasi suatu sentimen melalui proses preprocessing, bertujuan untuk memberikan rekomendasi kepada masyarakat tentang aplikasi dompet elektronik mana (GoPay, OVO, atau Dana) yang cocok untuk kegiatan jual beli sehari-hari. Studi ini menggunakan metode Naïve Bayes

Classifier untuk menganalisis sentimen di Twitter dan mengumpulkan komentar dan pendapat dari pengguna dalam bahasa Indonesia.

8. Pada penelitian yang dilakukan oleh (Alwan et al., 2022) ISSN: 2502-4752 dengan judul **“Diabetes Diagnosis System Using Modified Naive Bayes Classifier”** di dalam penelitian ini menyatakan Tiga tahap yang terlibat dalam sistem diagnosis diabetes yang diusulkan (DDS) yaitu pembuatan dataset, preprocessing, dan algoritma klasifikasi menggunakan Naïve Bayesian (TNB) tradisional dan Naïve Bayesian yang dimodifikasi (MNB). MNB Classifier adalah NB yang dimodifikasi yang digunakan untuk meningkatkan akurasi diagnosis dengan menambahkan model sederhana yang diusulkan untuk membantu memisahkan kelas diagnosis yang tumpang tindih. Hasil penelitian menunjukkan bahwa pengklasifikasi MNB umumnya lebih akurat daripada pengklasifikasi TNB untuk semua set fitur. Model TNB memiliki akurasi sekitar 63%, sedangkan model MNB memiliki akurasi (100%).

2.8 Kerangka Pemikiran

Kerangka pemikiran adalah alur logika berpikir yang menggabungkan antara teori atau konsep dengan berbagai variabel penelitian yang diidentifikasi sebagai masalah yang digambarkan dalam bentuk desain alur urutan penyelesaian penelitian. Berikut kerangka pemikiran yang dibuat oleh peneliti:



Gambar 2.2 Kerangka Penelitian

Sebagai input dalam penelitian ini adalah Data siswa-siswi SMP NEGERI 27 kota Batam. Setelah itu, data akan diproses menggunakan Metode Naive Bayes dan diimplementasikan dengan aplikasi data mining RapidMiner. untuk menghasilkan hasil prediksi tingkat pengaruh gadget terhadap pendidikan di SMP NEGERI 27 kota Batam.