Sentiment Analysis of Live.on Digital Provider Application Using Naive Bayes Classifier Method

Cendana Idli Mulia Asriguna, Tacbir Hendro Pudjiantoro, Puspita Nurul Sabrina and Asep Id Hadiana
Informatics Department, Faculty of Science and Informatics
Universitas Jenderal Achmad Yani, Cimahi, Indonesia
cendanaidlima18@if.unjani.ac.id, thp@if.unjani.ac.id, Puspita.sabrina@lecture.unjani.ac.id, asep.hadiana@lecture.unjani.ac.id

Abstract

Digital Provider In simple terms, this digital operator is an application-based prepaid card service that can be found on the Google Play Store and App Store. Live.On is a digital provider in Indonesia which has features such as freeing its users / users in choosing quota, topping quota, cellphone numbers can choose according to Live.On application recommendations, getting SIM cards and how to pay in just one application. Digital Provider users in Indonesia from the last year are growing rapidly. Sentiment Analysis identifies emotions and opinions both positive and negative. The Naïve Bayes method can be used to classify opinions into negative, and positive opinions. This research tries to take advantage by analyzing Google Play Store reviews and Indonesian-language reviews that talk about the Digital Provider brand. The stages of sentiment analysis in this study consist of retrieving data in the playstore using scrapping, pre-processing, lexicon based, naive bayes data classification, data evaluation and data visualization. Preprocessing is done by case folding, cleaning, tokenizing, and temming. The results obtained are the accuracy level obtained by the results of positive sentiment as much as 247 data, negative sentiment as much as 753. and using 80% training data and 20% test data from existing data. then the acquisition of results with classification using the naïve bayes method the accuracy value is 87%, recall is 81%, precision is 61% and F1 is 69%.

Keywords
Sentiment Analysis, Digital Provider, Naïve Bayes Classifier, Lexicon Based Feature, Live.On(10 font)

1. Introduction
Digital Provider In simple terms, this digital driver is an application-based prepaid card service that can be found on the Google Play Store or App Store. Other Digital Providers such as By.u, MPWR, PowerUp, and Live.On themselves have the same scope so it is necessary to research the extent of public interest, views on Digital Provider Live.On. The main difference is that with conventional prepaid services, all activities can be done digitally through applications or official websites.

Live.On is a digital provider in Indonesia from the XL Axiata company, Live.On was first created in Indonesia on August 17, 2020, but Live.On was only published on October 5, 2020 virtually where XL's 24th anniversary was taking place. Live.On here tries to make it easy for customers to make data package transactions. Live.On designed the application so that purchases can be made efficiently without having to meet. This also adapts to the potential when the market is in a COVID- 19 pandemic like today. where users are limited when transacting in person. Live.On has features such as freeing its users / customers in choosing quotas, beating quotas, cellphone numbers can choose according to recommendations from the Live.On application, getting simcards and payment methods in just one application, and cards will be sent for free / and can get die-commerce cards by using expedition services that already bear the shipping costs. So for Millennials it is very interesting Digital Providers like Live.On, and for the stoner interface system it is easy to understand so that novice users will not feel confused or difficult. Digital Provider users in Indonesia from year to year continue to grow rapidly.(A. R. Susanti 2016)

The aspects that drive satisfaction include product quality, price, service and ease of use or network. Application requests such as Google Playstore make it easy for users to get applications, and also the user review process can be done directly through comments provided by Google Playstore. Reviews on Google Playstore have a standing or

score from 1 to 5. However, often users who get the application provide a standing that does not match the review that should be so that this is not enough to describe the quality of the application through user reviews.(Putranti and Winarko 2014)People can simply communicate their emotions and feelings online by rating and reviewing products.via use of reviews and ratings. It will be necessary to study the idea of expressing thoughts and calculate insights to explore enterprises as a consequence of the rise in textual data.to investigate business, assess the idea of expressing feelings, and calculate insights.(Tripathy et al. 2015)Sentiment analysis on documents has been extensively studied, with Playstore being among the most well-known social media platforms where people may express their ideas on a variety of subjects in an objective manner (Susanti et al. 2017).

From the above background, a problem can be formulated as follows, in the research there are problems, namely there are many reviews of the digital provider application live.on with various kinds of sentiments from each review given by users of the application, a data collection is needed and specifically in this study, namely data labeling using lexicon based features to get the maximum accuracy value. Then calculate the confusion matrix which has the accuracy of Recall, Precision, F-Measure where this accuracy affects sentiment analysis in this study.

1.1 Objectives

Based on the factors identified, the researcher proposes to identify the sentiment of users of digital provider applications on google playstore reviews to analyze user opinions on Live.Onr knowing the accuracy level of the naïve bayes algorithm and lexicon based and can show the results of positive, negative tweets from each review. It is hoped that it can help users to find out sentiment info about the live.o digital provider application, and companies [un know the sentiment results from user reviews that have given opinion reviews of the live.on application. Price, service, speed, signal features are one of the considerations so that new users can find out what it is like, where to find out we do sentiment using lexicon labeling where each word has a negative or positive weight. This research also aims to determine the significant factors that affect the level of application deficiencies in terms of the following factors, service, price, signal speed mpbs, features in the application. proposes to generate sentiment based on factors and review data which will be preprocessed and then labeled using lexicon base d and naive bayes classification.

2. Literature Review

Naive Bayes classifier, a probabilistic classifier based on Bayes' theorem, is an excellent Bayesian network classifier. The Naive (Strong) independence assumption is considered by probabilistic classifiers built on Bayes' theorem. The problem of evaluating documents as belonging to one type or another by term frequency - inverted documents as features as features was first presented under various names into the text processing community and continues to be a common (basic) methodology for text categorization. The benefit of Nave Bayes is that it requires only a small quantity of training data to estimate the parameters associated with classification (Lopamudra D. et al. 2016).

Naive Bayes is a type of machine learning that does probability calculations using the Bayesian framework. In the Nave Bayes algorithm, Bayes' theorem is applied by fusing prior probability with conditional probability. To determine the likelihood of each potential categorization, an algorithm combines the prior probability and conditional probability in a formula.With a relatively straightforward structure and efficient structure, this naive bayes classifier algorithm is simple to develop and takes little processing time. Straight forward to create with a high-performing structure that is also very simple to use. The Nave Bayes classifier operates under the presumption that there is no significant relationship between the level or absence of a feature in a class and the existence of other features.

There is no correlation between the absence of one feature in a class and the existence of other features. The naive bayes classifier will continue operating on the assumption that features are independent and do not affect one another, despite the fact that features depend on one another (Yulita et al. 2021).

Opinion mining is the process of examining an author's viewpoint, feeling, or mindset from written text. the mindset of the creator of a written work. NLP, data mining, and machine learning are all used in opinion mining to do this task. Analyzing the specifications for opinion mining is the focus of this section. We focus on the sentiment data set and provide an outline of the review in the next section (Raju Shrestha 2019). Dictionary-based and corpus-based are the two sub-approaches for lexicon-based sentiment classification. Corpus-based. The dictionary-based strategy makes use of a dictionary and looks up highlighted terms there. A vast corpus is searched for opinion terms using a word list in the corpus-based technique in process of extracting semantic orientation.(V. D 2019)Companies can use this study to look at

client feedback and comments on the goods and services they offer. what they offer. Consequently, their future services will be improved. This will assist them in retaining current clients as well as attracting new ones in the future (Jabbar et al. 2018).

In sentiment analysis, there are three levels of classification: document-level classification, sentence-level classification and feature-level sentiment analysis. In document-level classification document, the main goal is to classify the opinions in theentire document as positive and negative (Bhavitha et al. 2017)This will not only help them to attract new customers in the future but also help them in retaining current customers.

The terms associated with the sense number are displayed in the Synsets Terms column. SentiwordNet's lexicon is extremely noisy; the majority of synsets have neither a positive nor a negative score. positive or unfavorable. The lexicon also overlooks lexical features that are pertinent to the material in the micro blog (Bonta et al. 2019). Confusion Matrix is used to perform the evaluation, which includes the True Positive Rate (TP Rate), True Negative Rate (TN Rate), False Positive Rate (FP Rate), and True False Negative Rate (TN Rate), False Positive Rate (FP Rate), and False Negative Rate (FN Rate) as indicators of False Negative Rate (FN Rate) (FN rate) (Tanesab et al. 2017).

Globally swift technical development in the trade, communications, and other industries is what defines the industry 4.0. One of them is the telecom sector. being the first digital operator in the nation and one of Telkomsel's digital service providers. Testing of the Support Vector Machine (SVM) algorithm using information from 300 application user reviews. There are 150 negative reviews for "by.u," many of which are negative or hateful. Additionally, 150 positive comments have been made, 150 of which contain links to constructive critique or honor the text's usage of Indonesian (Ardiansyah et al. 2020).

This research aims to analyze sentiment trends and measure the performance of the classification model so that itis expected to be a reference for business people to maintain the quality and development of by.U service providers and can be used as a reference for interested parties to develop by.U service providers. and the development of by.U service providers and can be used as a reference for parties or other researchers who have an interest in the case. or other researchers who have an interest in similar cases (Fransiska and Irham Gufroni 2020). Our sentiment lexicon is composed by words collected from from Twitter, as a representation of commonly used social media in Indonesia. commonly used social media in Indonesia. We build the lexicon by classifying the polarity of each word and enhancing it with some previously proven methods. some previously proven methods. The results of the tests and evaluations conducted in this study show that the evaluations conducted in this study show that In Set has a satisfactory performance as an Indonesian sentiment lexicon for predicting negative and Indonesian sentiment lexicon for predicting the negative and positive polarity of the written opinion negative and positive polarity of briefly written opinions (Koto and Rahmaningtyas 2018).

3. Methods

The researchers employed 1000 data that were obtained by web scraping from the Google Playstore in their complete investigation to process and choose the parameters to be taken into consideration in this study. and then split into two sentiment groups, positive and negative, with the former being done using a lexicon-based labeling end technique. The extracted data are labeled with sentiment VADER rules. Labeled data has undergone pre-processing such as tokenization, cleaning, and stemming. This research concentrates on reviews and user opinion of the digital service provider live.on, which is available here on the Google Play Store.
important indications and develop answers to crucial issues about the course of action. to make use of several algorithms, a combined approach combining Naive Bayes and Lexicon-Based Feature/Vader is suggested. The proposed methodology is fully illustrated in Figure 1 and may be discussed in four parts.

1. Data preprocessing is step one. Data must be cleaned and prepared for future analysis through data preparation. such as stemming, tokenizing, cleaning, and case folding.
2. Lexicon-Based Analysis: The data must be examined to obtain weights that are split into positive and negative categories. Step
3. Naive Bayes Analysis: To perform classification, two sets of data—training and test sets—must be prepared.
4. Determine accuracy using the confusion matrix, which also determines recall, precision, and f-measurement as well as accuracy of naive Bayes classification findings.

3.1 Data

Live Source Data The source of analysis for this study was reviews on the official Indonesian Playstore. This review data was created using phyton to produce scraping software. User reviews from the Live will be automatically retrieved. During the scraping process, on the app. The received review data will be categorized after the preprocessing stage is finished. The reviews will be divided into two types (categories) in this sentiment analysis system: positive sentiment class and negative sentiment class. This study needs both training and testing data, with the training data coming from a collection of reviews that have been manually categorized into sentiment groups. This data serves as training data for a sentiment analysis model. Then, using this model, categories will be created.reviews according to their emotions. Naive Bayes is the classification technique used in this research. We will use some scraping results as testing data. This test data uses a set of unlabeled reviews.

3.2 Pre Processing

a. Case Folding

During the case folding stage, all letters will be changed from uppercase to lowercase. One of the review examples's case folding procedures is as follows in Table 1:

Table 1. Pre Proccesing Case Folding

Before Case Folding	After Case Folding
Kenapa ya setia kali up date kuota TANGGAL MASA BERLAKUNYA SELALU DIKURANGI??? MISALNYA BELI KUOTA TGL 12 JAN, MASA BERLAKU CUMA SAMPAI 09 MARET, SO MASA BERLAKUNYA TIDAK 1 BULAN ?!..WHY?? TOLONG JELASKAN..SSLAIN ITU JUGA SUKA BANGET BUFFERING G SIGNAL LEMAH"	kenapa ya setia kali up date kuota tanggal masa berlakunya selalu dikurangi??? misalnya beli kuota tgl 12 jan, masa berlaku cuma sampai 09 maret, so masa berlakunya tidak 1 bulan ?!..why?? tolong jelaskan..sslain itu juga suka banget buffering g signal lemah"

b. Cleaning

The assessment team will proceed to the word removal stage after successfully completing the tokenizing phase. All of the review words will be examined. During the cleaning process, any pronouns or conjunctions that have nothing to do with sentiment analysis will be eliminated. The cleanup is as follows in Table 2:

Table 2. Pre Proccesing Cleaning

Before cleaning	After Cleaning
sinyal nya sering gangguan,kalo lagi gangguan bisa seharian g ada signal,padahal di kota,untuk paket data sangat murah banget...	sinyal nya sering gangguan kalo gangguan seharian g signal padahal kota paket data sangat murah banget

c. Tokenizing

The review will now be examined from the first character to the last, in reverse order. The i-th character will be concatenated with the next character if it lacks a word separator sign, such as a period (.), comma (,), space, or other separator symbol. The stages of the tokenizing stage are as follows in Table 3:

Table 3. Pre Proccesing Case Tokenizing

Before Tokenizing	After Tokenizing
sinyal nya sering gangguan kalo gangguan seharian g signal padahal kota paket data sangat murah banget	['sinyal', 'nya', 'sering', 'gangguan', 'kalo', 'gangguan', 'seharian', 'g', 'signal', 'padahal', 'kota', 'paket', 'data', 'sangat', 'murah', 'banget']

d. Stemming
Words in reviews frequently have numerous morphological variations. Each word is thereby condensed to an appropriate stemmed word (term). The words are taken in their most basic form by deleting prefixes and suffixes. general stemming stage procedures (Table 4).

Table 4.Pre Proccesing Stemming

Before Stemming	After Stemming
['sinyal', 'nya', 'sering', 'gangguan', 'jika', 'gangguan', 'seharian', 'g', 'sinyal', 'padahal', 'kota', 'paket', 'data', 'sangat', 'murah', 'benar-benar']	['sinyal', 'nya', 'sering', 'ganggu', 'jika', 'gangguan', 'seharian', 'g', 'sinyal', 'padahal', 'kota', 'paket', 'data', 'sangat', 'murah', 'sungguh']

3.3 Lexicon Based / Vader

The labeling strategy applied in this work is lexicon-based labeling. This technique uses lexicon features, which are dictionaries with both positive and negative weights. A list of terms having positive connotations and their weights are shown in Table X. In this language, any term that contains either positive or negative sentiment is weighted. The sentiment orientation of a word is determined by lexicon-based factors. One of these is the David Moeljadi Barasa Dictionary. To produce the SentiWordnet called Barasa, which is utilized in Indonesian, WordNet v1.1 and SentiWordNet 3.0 English were joined.

1. Calculate the total positive values and negative values in the text with the following formula

$$(1)\quad S_{positive} = \sum_{i=1}^{n} weight\ positivef_i$$

$$(2)\quad S_{negative} = \sum_{i=1}^{n} weight\ negativef_i$$

2. Calculate the compound value with the formula

$$b.\quad Compound_{positif} = \frac{S_{positif}}{\sqrt{S_{positif} * S_{positif} + \alpha}}$$

$$c.\quad Compound_{negatif} = \frac{S_{negatif}}{\sqrt{S_{negatif} * S_{negatif} + \alpha}}$$

3. Calculating the total value of $Compound_{positive}$ dan $Compound_{negative}$values

$$4.\quad Compound = Compound_{positif} + Compound_{negatif}$$

5. Determine the sentiment contained in the text with the rule that a sentence has a positivesentiment if it has a compond value ≥ 0, otherwise a sentence has a negative sentiment if the compond value < 0.

3.4. Naïve Bayes Classifier

The Naive Bayes technique develops a categorization model by employing Bayes' theorem. In this discussion, we will simulate the Naive Bayes strategy to classify text into positive or negative categorization. For the text classification simulation instance in this talk, the Naive Bayes algorithm is applied in the following order:

a. Determining the Term Frequency
b. Choosing Prior Probability
c. Using the formula to get likelihood probabilities

$$P(W_i|C)\frac{count(wiC) + 1}{count(C) + |v|}$$

d. Perform prediction on the test data with the formula

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

Table 5 represent the posterior likehood calculation.

Table 5. Posterior likelihood calculation

likehood Posterior	calculation	Probability Value
train 1		
P(sinyal lancar promo murah mantap sangat sukses terus | +) / dikalian seluruh hasil kata	0.067 * 0.067 * 0.067 * 0.134 * 0,5	6.64987E-07
P(sinyal lancar promo murah	0.09375 * 0.03125 * 0.03125	2.68221E-07

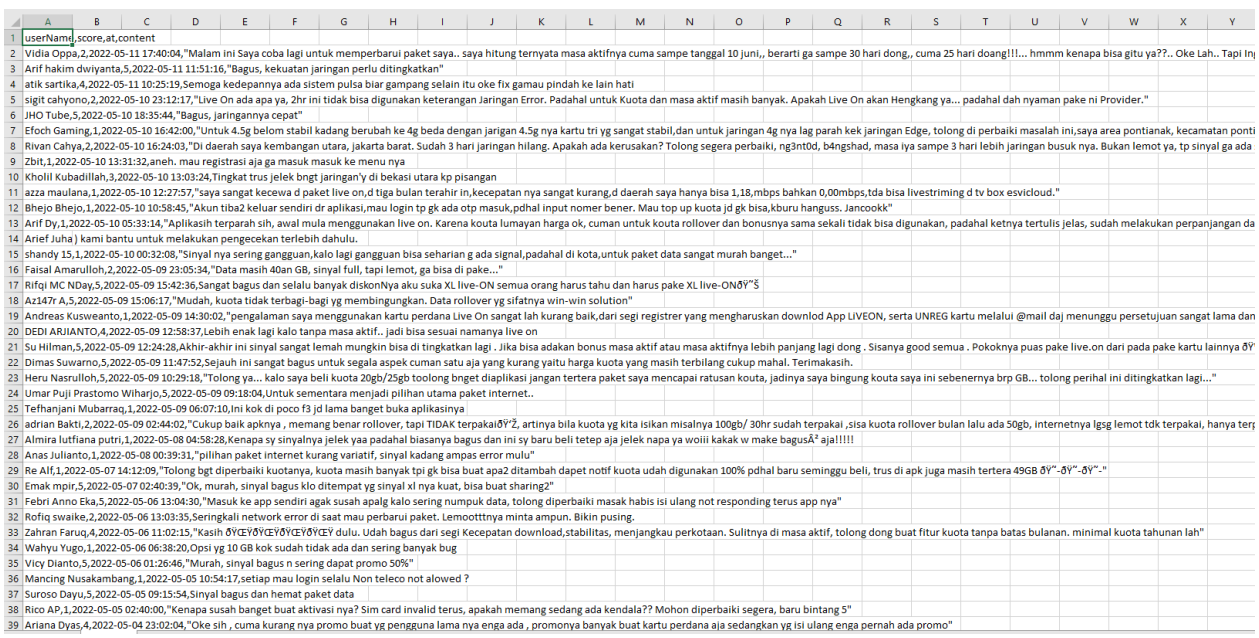
mantap sangat sukses terus | -) / dikalian seluruh hasil kata	* 0.0625 * 0,5	
train 2		
P(kecewa sinyal lemah pakai sudah sangat lama | +) dikalian seluruh hasil kata	0.033 * 0.067 * 0.033 * 0.033 * 0.033 * 0,5	3.97284E-08
P(kecewa sinyal lemah pakai sudah sangat lama | -) dikalian seluruh hasil kata	0.0625 * 0.09375 * 0.09375 * 0.0625 * 0.09375 * 0,5	1.60933E-06

4. Data Collection

Scrapping live.on review data from the Google Play Store site is done using the Python programming language by utilizing the google-play-scrapper library with 1000 review data taken Data is filtered according to the data needed. The data used in this study is the latest data from May 05, 2022. The data is then labeled with sentiment. Labeling process. The labeling process is carried out using Lexicon based or vader lexicon to determine the weight to determine which reviews are more positive or negative based on word usage or the number of positive and negative words to get positive and negative results.

- Unlabeled

Raw data scrapping results that have not been labeled with username, date, review score, and review text (Figure 1).



```
userName,score,at,content
Vidia Oppa,2,2022-05-11 17:40:04,"Malam ini Saya coba lagi untuk memperbarui paket saya.. saya hitung ternyata masa aktifnya cuma sampe tanggal 10 juni,, berarti ga sampe 30 hari dong,, cuma 25 hari doang!!!... hmmm kenapa bisa gitu ya??.. Oke Lah.. Tapi In
Arif hakim dwiyanta,5,2022-05-11 11:51:16,"Bagus, kekuatan jaringan perlu ditingkatkan"
atik sartika,4,2022-05-11 10:25:19,Semoga kedepannya ada sistem pulsa biar gampang selain itu oke fix gamau pindah ke lain hati
sigit cahyono,2,2022-05-10 23:12:17,"Live On ada apa ya, 2hr ini tidak bisa digunakan keterangan Jaringan Error. Padahal untuk Kuota dan masa aktif masih banyak. Apakah Live On akan Hengkang ya... padahal dah nyaman pake ni Provider."
JHO Tube,5,2022-05-10 18:35:44,"Bagus, jaringannya cepat"
Efoch Gaming,1,2022-05-10 16:42:00,"Untuk 4.5g belom stabil kadang berubah ke 4g beda dengan jarigan 4.5g nya kartu tri yg sangat stabil,dan untuk jaringan 4g nya lag parah kek jaringan Edge, tolong di perbaiki masalah ini,saya area pontianak, kecamatan ponti
Rivan Cahya,2,2022-05-10 16:24:03,"Di daerah saya kembangan utara, jakarta barat. Sudah 3 hari jaringan hilang. Apakah ada kerusakan? Tolong segera perbaiki, ng3nt0d, b4ngshad, masa iya sampe 3 hari lebih jaringan busuk nya. Bukan lemot ya, tp sinyal ga ada
Zbit,1,2022-05-10 13:31:32,aneh. mau registrasi aja ga masuk masuk ke menu nya
Kholil Kubadillah,3,2022-05-10 13:03:24,Tingkat trus jelek bngt jaringan'y di bekasi utara kp pisangan
azza maulana,1,2022-05-10 12:27:57,"saya sangat kecewa d paket live on,d tiga bulan terahir in,kecepatan nya sangat kurang,d daerah saya hanya bisa 1,18,mbps bahkan 0,00mbps,tda bisa livestriming d tv box esvicloud."
Bhejo Bhejo,1,2022-05-10 10:58:45,"Akun tiba2 keluar sendiri dr aplikasi,mau login tp gk ada otp masuk,pdhal input nomer bener. Mau top up kuota jd gk bisa,kburu hanguss. Jancookk"
Arif Dy,1,2022-05-10 05:33:14,"Aplikasih terparah sih, awal mula menggunakan live on. Karena kouta lumayan harga ok, cuman untuk kouta rollover dan bonusnya sama sekali tidak bisa digunakan, padahal ketnya tertulis jelas, sudah melakukan perpanjangan da
Arief Juha) kami bantu untuk melakukan pengecekan terlebih dahulu.
shandy 15,1,2022-05-10 00:32:08,"Sinyal nya sering gangguan,kalo lagi gangguan bisa seharian g ada signal,padahal di kota,untuk paket data sangat murah banget..."
Faisal Amarulloh,2,2022-05-09 23:05:34,"Data masih 40an GB, sinyal full, tapi lemot, ga bisa di pake..."
Rifqi MC NDay,5,2022-05-09 15:42:36,Sangat bagus dan selalu banyak diskonNya aku suka XL live-ON semua orang harus tahu dan harus pake XL live-ONðŸ"Š
Az147r A,5,2022-05-09 15:06:17,"Mudah, kuota tidak terbagi-bagi yg membingungkan. Data rollover yg sifatnya win-win solution"
Andreas Kusweanto,1,2022-05-09 14:30:02,"pengalaman saya menggunakan kartu perdana Live On sangat lah kurang baik,dari segi registrer yang mengharuskan downlod App LiVEON, serta UNREG kartu melalui @mail daj menunggu persetujuan sangat lama dan
DEDI ARJANTO,4,2022-05-09 12:58:37,Lebih enak lagi kalo tanpa masa aktif.. jadi bisa sesuai namanya live on
Su Hilman,5,2022-05-09 12:24:28,Akhir-akhir ini sinyal sangat lemah mungkin bisa di tingkatkan lagi . Jika bisa adakan bonus masa aktif atau masa aktifnya lebih panjang lagi dong . Sisanya good semua . Pokoknya puas pake live.on dari pada pake kartu lainnya ðŸ
Dimas Suwarno,5,2022-05-09 11:47:52,Sejauh ini sangat bagus untuk segala aspek cuman satu aja yang kurang yaitu harga kuota yang masih terbilang cukup mahal. Terimakasih.
Heru Nasrulloh,5,2022-05-09 10:29:18,"Tolong ya... kalo saya beli kuota 20gb/25gb toolong bnget diaplikasi jangan tertera paket saya mencapai ratusan kouta, jadinya saya bingung kouta saya ini sebenernya brp GB... tolong perihal ini ditingkatkan lagi..."
Umar Puji Prastomo Wiharjo,5,2022-05-09 09:18:04,Untuk sementara menjadi pilihan utama paket internet..
Tefhanjani Mubarraq,1,2022-05-09 06:07:10,Ini kok di poco f3 jd lama banget buka aplikasinya
adrian Bakti,2,2022-05-09 02:44:02,"Cukup baik apknya , memang benar rollover, tapi TIDAK terpakaiðŸ"Ž, artinya bila kuota yg kita isikan misalnya 100gb/ 30hr sudah terpakai ,sisa kuota rollover bulan lalu ada 50gb, internetnya lgsg lemot tdk terpakai, hanya terp
Almira lutfiana putri,1,2022-05-08 04:58:28,Kenapa sy sinyalnya jelek yaa padahal biasanya bagus dan ini sy baru beli tetep aja jelek napa ya woiii kakak w make bagusÂ² aja!!!!!
Anas Julianto,1,2022-05-08 00:39:31,"pilihan paket internet kurang variatif, sinyal kadang ampas error mulu"
Re Alf,1,2022-05-07 14:12:09,"Tolong bgt diperbaiki kuotanya, kuota masih banyak tpi gk bisa buat apa2 ditambah dapet notif kuota udah digunakan 100% pdhal baru seminggu beli, trus di apk juga masih tertera 49GB ðŸ"˜-ðŸ"˜-ðŸ"˜-"
Emak mpir,5,2022-05-07 02:40:39,"Ok, murah, sinyal bagus klo ditempat yg sinyal xl nya kuat, bisa buat sharing2"
Febri Anno Eka,5,2022-05-06 13:04:30,"Masuk ke app sendiri agak susah apalg kalo sering numpuk data, tolong diperbaiki masak habis isi ulang not responding terus app nya"
Rofiq swaike,2,2022-05-06 13:03:35,Seringkali network error di saat mau perbarui paket. Lemootttnya minta ampun. Bikin pusing.
Zahran Faruq,4,2022-05-06 11:02:15,"Kasih ðŸŒŸðŸŒŸðŸŒŸðŸŒŸ dulu. Udah bagus dari segi Kecepatan download,stabilitas, menjangkau perkotaan. Sulitnya di masa aktif, tolong dong buat fitur kuota tanpa batas bulanan. minimal kuota tahunan lah"
Wahyu Yugo,1,2022-05-06 06:38:20,Opsi yg 10 GB kok sudah tidak ada dan sering banyak bug
Vicy Dianto,5,2022-05-06 01:26:46,"Murah, sinyal bagus n sering dapat promo 50%"
Mancing Nusakambang,1,2022-05-05 10:54:17,setiap mau login selalu Non teleco not alowed ?
Suroso Dayu,5,2022-05-05 09:15:54,Sinyal bagus dan hemat paket data
Rico AP,1,2022-05-05 02:40:00,"Kenapa susah banget buat aktivasi nya? Sim card invalid terus, apakah memang sedang ada kendala?? Mohon diperbaiki segera, baru bintang 5"
Ariana Dyas,4,2022-05-04 23:02:04,"Oke sih , cuma kurang nya promo buat yg pengguna lama nya enga ada , promonya banyak buat kartu perdana aja sedangkan yg isi ulang enga pernah ada promo"
```

Figure 1. Data Unlabeled

- Labeled

Data that has been labeled as a result of pre-processing and the results of the lexicon-based feature method that has produced negative and positive results on each review (Figure 2).

	A	B	
1	no	content	label
2	1	malam coba baru paket hitung nyata masa aktif cuma sampai tanggal juni arti tidak sampai hari dong cuma hari doang hmmm begitu oke lah ingat kalo signal live masih lot terus data tidak dipakai sampai tanggal juni fixx bulan depan saya balik pakai indosat biarp	Negatif
3	2	bagus kuat jaring perlu tingkat	Positif
4	3	moga depan sistem pulsa biar gampang itu oke fix tidak mau pindah lain hati	Negatif
5	4	live ada apa hari tidak guna terang jaring error padahal kuota masa aktif banyak live akan hengkang padahal deh nyaman pakai ini provider	Negatif
6	5	bagus jaring cepat	Positif
7	6	tidak bom stabil kadang ubah tidak beda jarigan tidak kartu tri yang sangat stabil untuk jaring tidak lag parah kayak jaring edge di baik masalah area pontianak camat pontianak barat moga keluh saya di dengar tim kait terima kasih	Negatif
8	7	daerah kembang utara jakarta barat hari jaring hilang rusa segera baik ng nanti ada b ngshad masa iya sampai hari lebih jaring busuk bukan lot tapi tidak se bar gue menyesal beli kartu bom ada bulan lanjur rugi gue segera baik b angshad ternyta live lebih b saja ng	Negatif
9	8	aneh mau registrasi saja tidak masuk masuk menu	Negatif
10	9	tingkat terus jelek banget jaring ya bekas utara kp pisang	Positif
11	10	sangat kecewa ada paket live ada tiga bulan terahir ini cepat sangat kurang ada daerah hanya mbps bahkan mbps tda livestriming ada tv box esvicloud	Negatif
12	11	akun tiba keluar sendiri dari aplikasi mau login tapi tidak otp masuk padahal input nomer benar mau top up jadi tidak kburu hanguss jancookk	Negatif
13	12	aplikasih parah sih awal mula guna live kouta lumayan harga cuman kouta rollover bonus sama sekali guna padahal ketnya tulis jelas laku panjang beli kouta tambah aplikasi tiba tiba tidak bisa di buka alias blank putih tidak jelas ampas lah ini	Negatif
14	13	belum mohon maaf buat kamu repot hubung tahu detail yang milik ini bantu kece informasi mohon tunggu kak kami bantu untuk laku kece lebih	Negatif
15	14	nya sering ganggu kalo ganggu hari tidak signal padahal kota paket data sangat murah banget	Negatif
16	15	data an full lot tidak di pakai	Negatif
17	16	sangat bagus selalu banyak diskon aku suka live semua orang tahu harus pakai live	Negatif
18	17	mudah tidak bagi yang bingung data rollover yang sifat win win solution	Negatif
19	18	alam guna kartu perdana live sangat lah kurang baik segi registrer harus downlod app liveon unreg kartu lalu daj tunggu tuju sangat lama ribet paling parah quota sisa dengan jatuh tempo hari masa panjang paket bisa gun maksud apa mengajak ribut padahal kali j	Negatif
20	19	lebih enak kalo masa aktif bisa sesuai nama live	Positif
21	20	akhir akhir sinyal sangat lemah mungkin di tingkat bisa adakan bonus masa aktif masa aktif lebih panjang dong sisa bagus semua pokok puas pakai live dari pakai kartu lain	Negatif
22	21	jauh sangat bagus segala aspek cuman satu saja kurang harga yang bilang cukup mahal terima kasih	Negatif
23	22	ya kalo beli toolong banget aplikasi jangan tera paket capai ratus kouta jadi bingung kouta sebenarnya berapa gb perihal ini tingkat	Positif
24	23	sementara jadi pilih utama paket internet	Negatif
25	24	kok poco f jadi lama banget buka aplikasi	Negatif
26	25	cukup baik apknya memang benar rollover tidak pakai arti bila yang isi misal hari pakai sisa rollover bulan lalu gb internetnya langsung lot tidak pakai pakai akses apk untuk beli baru	Negatif
27	26	saya sinyal jelek ya padahal biasa bagus ini saya baru beli tetap saja jelek kenapa woiii kakak gue memakai bagus saja	Negatif
28	27	pilih paket internet kurang variatif kadang ampas error mulu	Negatif
29	28	banget baik kuota masih banyak tapi tidak buat apa tambah dapat notif sudah guna padahal baru minggu beli terus apk masih tera	Negatif
30	29	murah bagus kalo tempat yang xl kuat buat sharing	Negatif
31	30	masuk app sendiri susah apalagi kalo sering numpuk data baik masak habis isi ulang not responding terus app	Negatif
32	31	seringkali network error saat mau baru paket lemoottttnya meminta ampun bikin pusing	Negatif
33	32	kasih dulu sudah bagus segi cepat download stabilitas jangkau kota sulit masa aktif dong buat fitur tanpa batas bulan minimal tahun lah	Negatif
34	33	opsi yang kok tidak dan sering banyak bug	Negatif
35	34	murah bagus n sering promo	Positif
36	35	mau login selalu non teleco not alowed	Positif
37	36	bagus hemat paket data	Negatif
38	37	susah banget buat aktivasi sim card invalid terus memang sedang kendala mohon baik segera baru bintang	Negatif
39	38	oke sih cuma kurang promo buat yang guna lama enga promonya banyak buat kartu perdana saja yang isi ulang enga pernah promo	Negatif

dataset_xl_3

Figure 2. Data Labeled

5. Graphical Results

In order to determine the outcomes with the utmost precision, the model's performance is now assessed by testing the data with each class. The percentage of the total data that is correctly recognized and assessed is measured when utilizing Confusion Matrix (Figure 3), which takes the shape of a matrix table (Table 6).

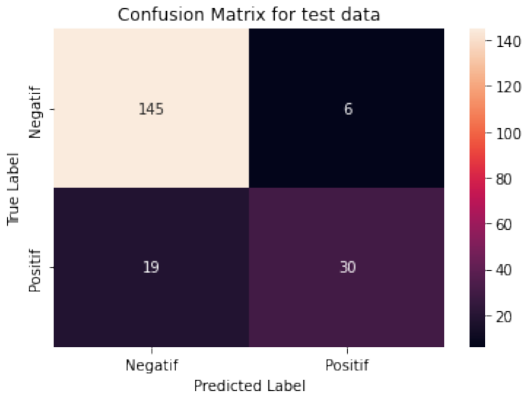


Figure 3. Confusion Matrix

Table 6. Confusion Matriks

	Aktual +	Aktual -
Prediksi +	TP = 30	FP = 19
Prediksi -	FN = 7	TN = 144

a. Preccision
A precise computation is made by dividing the number of true positive data by the sum of true positive and false positive data.
- presisi: (TP + FP)/TP

b. Recall
Recall is the ratio of the relevant content that has been discovered to something which actually exists. The recall computation is carried out by multiplying the true positive data by the sum of true positive and false

negative data. The false negative data value is determined by counting all values other than the actual positive row for each class.
Recall: TP/(TP+FN)

c. F-Measurement
F-measure A special retrieval success parameter called F-measure combines recall and precision. The F-measure value is calculated by multiplying the product of recall and precision by the sum of recall and precision, then dividing by the result. Accuracy is defined as the percentage of all sentiments that are correctly detected. Accuracy is computed by dividing the number data and test data by the number of true positive sentiment data.
F-Measure is equal to 2*(Precision*Recall) / (Precision+Recall).

d. Accuracy
Accuracy is defined as the percentage of the total feelings that are correctly detected. The accuracy calculation is determined by splitting the total data and test data by the number of correct sentiment data.
Accuracy is equal to (TP+TN)/(TP+TN+FP+FN)

6. Conclusion

Based on the results of the final project research, it can be concluded that the sentiment of users of the live.on digital provider application is more negative reviews than positive reviews where the use of lexicons helps to label positive and negative classes so that the process and results of the naïve bayes classification model in this study get an accuracy value (87%). The data used is 1000 datasets which are taken directly using the scrapping technique on Google playstore. With a recall test accuracy rate of 81%, presicion 61%, f-measurement 69% shows that the test accuracy results are greater using recall.

References

Susanti, A. R., Analisis Klasifikasi Sentimen Twitter Terhadap Kinerja Layanan Provider Telekomunikasi Menggunakan Varian Naïve Baye, *Institut Pertanian Bogor*, 2016.

Alkubaisi, A. J., Abdulsattar, G., Kamaruddin, S. S. and Husni, H., Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers, *Computer and Information Science* vol. 11, no. 1, pp. 52, 2018.

Ardiansyah, A., Sopian A., Dany, P., Sandra, J. K., Octa, P. P., and Cep, A., The Analysis of Digital Provider Sentiment of 'by.u' on Google Play Which Uses the Support Vector Machine (SVM) Method, *Journal of Physics: Conference Series,* vol. 1641, no. 1, 2020.

Bhavitha, B. K., Anisha P. R., and Niranjan N. C., Comparative Study of Machine Learning Techniques in Sentimental Analysis, *Proceedings of the International Conference on Inventive Communication and Computational Technologies, ICICCT 2017* (October), pp. 216–21, 2017.

Bonta, V., Nandhini, K. and Janardhan. N., A Comprehensive Study on Lexicon Based Approaches for Sentiment Analysis." *Asian Journal of Computer Science and Technology,* vol. 8, no. S2, pp. 1–6, 2019.

Chaithra, V. D., Hybrid Approach: Naive Bayes and Sentiment VADER for Analyzing Sentiment of Mobile Unboxing Video Comments, *International Journal of Electrical and Computer Engineering (IJECE),* vol. 9, no. 5, pp. 4452, 2019.

Lopamudra Dey, Sanjay Chakraborty, Anuraag Biswas, Beepa Bose, and Sweta Tiwari. 2016. "Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier." *International Journal of Information Engineering and Electronic Business* 8(4):54–62. doi: 10.5815/ijieeb.2016.04.07.

Fransiska, S. and Gufroni, A. C., Sentiment Analysis Provider by.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method, *Scientific Journal of Informatics,* vol. 7, no. 2, pp. 2407–7658, 2020.

Fajri, K. and Rahmaningtyas, G. Y., Inset Lexicon: Evaluation of a Word List for Indonesian Sentiment Analysis in Microblogs, *Proceedings of the 2017 International Conference on Asian Language Processing, IALP 2017* 2018-January(December), pp. 391–94, 2018.

Putranti, N. D. and Winarko, E., Analisis Sentimen Twitter Untuk Teks Berbahasa Indonesia Dengan Maximum Entropy Dan Support Vector Machine, *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 8, no. 1, pp. 91, 2014.

Shrestha, R., Sentiment Analysis of Twitter Data Using Logistic Regression, vol. 3, no. 6, pp. 144–47, 2019.

Aisah Rini, S., Djatna, T. and Kusuma, W. A., Twitter's Sentiment Analysis on GSM Services Using Multinomial Naïve Bayes, *Telkomnika (Telecommunication Computing Electronics and Control),* vol. 15, no. 3, pp. 1354–61, 2017.

Tanesab, F. I., Irwan S., and Hindriyanto, D. P., Sentiment Analysis Model Based On Youtube Comment Using Support Vector Machine, *International Journal of Computer Science and Software Engineering (IJCSSE),* vol. 6, no. 8, pp. 180–85, 2017.

Tripathy, A., Ankit A. and Rath, S. K., Classification of Sentimental Reviews Using Machine Learning Techniques, *Procedia Computer Science,* vol. 57, pp. 821–29, 2015.

Yulita, W., Eko, D. N. and Algifari, M. H., Program Studi Teknik Informatika, Institut Teknologi Sumatera, Jl Terusan Ryacudu, Way Huwi, Jati Agung, and Lampung Selatan. 2021, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *Jdmsi,* vol. 2, no. 2, pp. 1–9, 2021.

Biographies

Cendana Idli Mulia Asriguna is a final year undergraduate student in the department of informatics, University Jenderal Achmad Yani, Cimahi, Indonesia. Her primary interests are Data Mining, Systems Analysis, and software engineering.

Tacbir Hendro Pudjiantoro is an Associate Professor. Doctoral Candidate from the Indonesian University of Education. Researcher in the field of Knowledge Management and handles several Information Systems projects.

Puspita Nurul Sabrina is an Associate Lecturer. Received her Master's degree in Informatics from Insitute Techonology Of Bandung. Amongst her Researcher interest are Data mining, Software Quality, Business Intelegent.

Asep Id Hadiana received his Master's degree in Enterprise Information System from Indonesian Computer Univerity and a Doctor of Philosophy from Technical University of Malaysia Melaka. He is a lecturer in the Informatics Department, Faculty of Science and Informatics, Universitas Jenderal Achmad Yani. Amongst his research interest are Cyber Security, Data Mining, Spatial Analysis, Location Based Services and Geographic Information Systems.