|  |  |
| --- | --- |
| SLIDE 1 | Assalamualaikum Warahmatullahi Wabarakatuh/ I’m Imam Suyuti/ from Universitas Sebelas Maret/ Indonesia// I would like to present my article and title “Fine-Grained Sentiment Analysis/ on PeduliLindungi Application Users/ with Multinomial Naïve Bayes-SMOTE”// |
| SLIDE 2 | Increasing spread of COVID-19 virus in Indonesia has prompted the government to issue a policy to control the spread of the virus/ by imposing large-scale restrictions/ which are often known as Large-Scale Social Restrictions//  PT Telekomunikasi Indonesia/ and Ministry of Communication and Information have collaborated to create the PeduliLindungi application// This application was developed by the government/ to assist in tracking /and stopping the spread of Covid-19 virus//  In each application/ there are user ratings and reviews regarding the services/ and features provided// Reviews can be in the form of suggestions/ criticisms/ or complaints// |
| SLIDE 3 | **Sentiment analysis** is computational research of textually expressed **opinions/** **sentiments/** and **emotions//** In this case/ sentiment analysis can be applied to **mobile application reviews//**  The problem that often arises in most sentiment analysis studies/ is that most of the review data tend to be unbalanced/ in terms of the number of each class//  In this research/ the Synthetic Minority Oversampling Technique (**SMOTE**) method/ was used to handle the case of unbalanced data// |
| SLIDE 4 | I use five articles on sentiment analysis/ as a reference in creating my articles// |
| SLIDE 5 | **Primary data** was collected through a web scrapping process/ on the Google Play Store website/ and data was collected from **September 14** to **October 14/ 2021/** with as many as **9021** PeduliLindungi user reviews//  **Web Scrapping method** is a method used to collect/ or extract **semi-structured information/** or **data** from websites//  Then/ the process of analyzing/ and exploring unstructured text data/ becomes structured to identify patterns in the form of new knowledge/ and meaningful or useful information/ with the help of a technology is called **text mining//** |
| SLIDE 6 | **Text preprocessing proses** is a prepared document text/ or raw data set proses to convert text data//  The **steps for text processing** are as follows:  **Case Folding** /  **Tokenizing** /  **Filtering** /  **and Stemming** //  **TF-IDF weighting** is one of the models used to calculate word weights/ in a document using the term frequency model to count the words that appear in each document// |
| SLIDE 7 | In this research/ **Fine-Grained Sentiment Analysis** is used which is often referred to as aspect-based opinion mining/ and its basic tasks include **aspect extraction/** **opinion identification/** and **sentiment classification//**  This analysis includes **positive** and **negative sentiment//**  **VADER Lexicon** is used to automatically label sentiment classes on English review data//  Data that has been labeled sentiment class will then be partitioned into training and testing/ with **80:20** data split//  **Synthetic Minority Oversampling Technique** (**SMOTE**) method/ is used to handle the case of **unbalanced data//** |
| SLIDE 8 | learning process is carried out using the supervised learning classification method/ namely **Multinomial Naive Bayes//**  Multinomial Naive Bayes is a supervised learning method that uses **probability/** and is more focused on **text classification//**  The calculation of the probability of review *d* having class *c* can be seen in the following/ probability of *a* class *c* on document/text *d* proportional to prior probability c/ pi/ probability of a word in class c//  The calculation of probability of positive sentiment or negative sentiment from a review is/ number of occurrences of positive or negative sentiment words in the text/ division by/ number of occurrences of positive and negative sentiment words in the text// |
| SLIDE 9 | number of each prediction class/ and actual class must be known when measuring the accuracy of the classification/ following Table **confusion matrix** can be used to calculate classification accuracy//  If there is a **data imbalance case/** then choosing which model is best/ it can be done by using the **AUC/** (**Area Under Curve**) value// |
| SLIDE 10 | **AUC** is an evaluation criterion that uses sensitivity/ or specificity as the basis for measurement// **AUC value** is generally in the interval **0.5 to 1.0**// following table description for each AUC value interval can be seen in Table/  Wordcloud will be used to **visualize the results**/ Wordcloud is a representation of data that shows a set of important/ and frequently occurring words in a word// |
| SLIDE 11 | RESEARCH RESULT |
| SLIDE 12 | **Text preprocessing stage** is needed to uniform the spelling of words and letter shapes/ reduce vocabulary/ and eliminate characters other than letters/ so that the data text is more structured/ and informative// The results of text preprocessing are in Table below// |
| SLIDE 13 | This research uses the **VADER lexicon library/** to carry out the process of labeling sentiment categories/ on English review data automatically//  The review data will only be labeled as two types of sentiment/ **positive sentiment/** and **negative sentiment//** from the table/ there is 2777 positive sentiment/ and 6244 negative sentiment// |
| SLIDE 14 | Based on the review text/ that will be analyzed and has passed the preprocessing stage/ formula used to calculate the sentiment score is/ positive words count minus by negative wors count  The simulation results of the sentiment score calculation can be seen as follows/ From review text/ it have 1 positive words and 2 negative words/ so the sentiment label is negative// |
| SLIDE 15 | This research will use **SMOTE** and **non-SMOTE** techniques/ Classification predictions are made with a **Multinomial Naïve Bayes** algorithm// From the results of the classification method used the **AUC/** **recall/** **precision/** and **accuracy values//**  This analysis will partition the dataset with a ratio of **80%** as training data/ and **20%** as test data//  From the table/ Multinomial naïve bayes with smote has the higher AUC value than non-smote// |
| SLIDE 16 | Wordcloud will be used to **visualize the results** of the classification analysis// The purpose of the visualization is **to extract information in the form of topics/** that are often discussed by PeduliLindungi users//  From pedulilindungi negative sentiment worcloud/ the most word is difficult, bad, error, fix, slow, time, and problem/ and the most word from pedulilindungi positive sentiment wordcloud is care, protect, good, help,  better, and thank// |
| SLIDE 17 | **Last, Conclution// SMOTE** is proven to increase the accuracy of the model on unbalanced data seen from the AUC value in the model/ with SMOTE higher than the AUC value generated by the non-SMOTE model// The resulting AUC value is **0.9061/** so the method is classified as having a **very good classification performance/** and user reviews have more **negative sentiments** related to the **error application system/** while **positive reviews** relating to the **benefits of using the application//** |
| SLIDE 18 | THANKS |