SAGA: Journal of Technology and Information Systems
Vol 2, Issue 3, August 2024, Page 288-294
ISSN: 2985-8933 (Media Online)
DOI: 10.58905/SAGA.v2i3.334

Sentiment Analysis on Erspo Jersey in X Using Machine Learning Algorithms

Andi Asrida Reskinah. D[1], Marhawati Najib[2], Muhammad Ashdaq[3]

[1]Department of Economic Education, Universitas Negeri Makassar, Makassar, Indonesia
[2,3]Department of Digital Business, Universitas Negeri Makassar, Makassar, Indonesia

Author Email: marhawati@unm.ac.id[1], andi.rdhaa18@gmail.com[2], muhammad_ashdaq@unm.ac.id[3]

Abstract. This research conducts a sentiment analysis on Erspo jerseys using machine learning algorithms on the X platform. The objective is to identify the public's sentiment and compare the performance of three algorithms: Naïve Bayes, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM). Data was collected through web scraping of tweets between January and September 2024, containing keywords related to Erspo. Using a lexicon-based approach, the preprocessing steps involved cleaning, tokenizing, normalizing, and labeling data into positive, negative, and neutral sentiments. Results show that the Naïve Bayes algorithm provided the highest accuracy in sentiment classification, followed by SVM and KNN. Positive sentiment primarily centered on product loyalty, while negative sentiment largely criticized jersey design and quality. The findings offer important insights for Erspo stakeholders to refine marketing strategies and product improvements. This study highlights the potential of machine learning in analyzing consumer opinions at scale, making it a valuable tool for real-time consumer feedback analysis.

Keywords: Sentiment Analysis, Machine Learning, Erspo Jersey, X

1 Introduction

In recent years, the sports industry has experienced rapid development not only in terms of athlete achievements but also in the commercialization and marketing of sports products. One of the most popular products among sports fans is the jersey or team uniform. Jerseys not only serve as a team identity but also as a symbol of pride for fans.

Erspo, a local brand specializing in jersey production, has successfully attracted attention with its competitive product designs and quality. Its presence in various local markets, including X, shows its significant potential in Indonesia's sports apparel industry. However, amidst the increasingly fierce competition with other brands, Erspo faces various challenges in maintaining its position as the top choice for consumers.

Local companies such as Erspo have sought to capitalize on opportunities in the sports apparel market by offering competitive jerseys. However, the brand faces challenges in maintaining customer loyalty, especially regarding product quality controversies that have affected consumer sentiment, particularly on social media platforms like X (formerly Twitter). According to a report from We Are Social, Indonesia ranks fourth globally in the number of active X users, making this platform a crucial source for understanding public opinion about brands. As of October 2023, there are 666.2 million X users worldwide, with the number increasing by 18.1% quarter-to-quarter (qtq) and 22.4% year-on-year (yoy). X has thus become a key platform where consumers freely express opinions and share experiences, providing essential feedback for brands to gauge how their products are perceived in the market. The sentiment on X can reflect customer satisfaction, quality perception, and reactions to new jersey designs by Erspo. Therefore, it is crucial for businesses to conduct sentiment analysis to help them boost sales based on unique and real customer feedback[1].

Various studies have emphasized the significance of sentiment analysis in helping companies understand public perceptions of their products. For instance, some research explored the effectiveness of algorithms like Naïve Bayes for sentiment analysis on social media platforms[2][3]. Meanwhile, others introduced methodologies such as SEMMA to enhance sentiment analysis in the context of e-commerce. These methods have proven valuable in improving the accuracy and insightfulness of sentiment analysis, enabling businesses to better respond to consumer feedback and market trends[4][5].

However, these studies have been limited to specific aspects of certain products or services and have not specifically examined sentiment analysis for local Indonesian brands like Erspo. Hence, this research aims to fill that gap by conducting sentiment analysis on Erspo jerseys on the X platform using machine learning algorithms such as Naïve Bayes, K-Nearest Neighbor, and Support Vector Machine.

Sentiment analysis involves the automatic extraction, understanding, and processing of information from unstructured text to obtain sentiment data from statements or opinions[6]. Sentiment analysis can be categorized into positive, negative, and neutral sentiments. However, it is generally better to classify sentiment into positive and negative categories, as neutral sentiment is considered to carry little value[7].

Traditionally, sentiment analysis has been done manually, requiring the labor-intensive task of reading and evaluating thousands or millions of comments. This manual approach is not only time-consuming and costly but is also prone to bias and human error. The use of machine learning (ML) technology in sentiment analysis is thus essential.

By using machine learning algorithms, companies can analyze sentiment faster, more accurately, and on a larger scale. Various classification algorithms can be used, including Support Vector Machine, K-Nearest Neighbor, and Naïve Bayes[8]. This research differs from previous studies as it focuses on a local Indonesian brand, Erspo, which faces reputation challenges in the jersey market. Additionally, this study employs a combination of several machine learning algorithms (Naïve Bayes, K-Nearest Neighbor, and Support Vector Machine), which have not been widely applied together to similar research objects in Indonesia. This approach is expected to provide more comprehensive and accurate results in understanding public sentiment towards Erspo’s products.

In this study, sentiment analysis on Erspo jerseys on the X platform will be performed by comparing machine learning algorithms namely, Naïve Bayes, K-Nearest Neighbor, and Support Vector Machine under the research topic titled *"Sentiment Analysis on the Erspo Jersey on X Using Machine Learning Algorithms."*

2 Methodology

This study uses qualitative research. States that qualitative research is research that intends to interpret phenomena experienced by research subjects such as behavior, perception, motivation, actions, and others holistically and by means of description in the form of words and language, in a specific context that is natural and by utilizing various natural methods[9]. States that descriptive research is research that is intended to investigate circumstances, conditions, or other things. This study aims to find out and find as much information as possible and provide an overview related to Jersey Erspo[10].

The place of research in this study was conducted online on the X platform as a source of publicly available data on the X social media platform, from June to August 2024. The population in this study includes all tweets or messages shared by Twitter users related to Jersey Erspo. While the data used as a sample is data related to tweets taken in the last 7 months that are relevant to this study.

This research begins with the data collection stage through the Twitter API. At the sample stage, the researcher performs data crawling on the data to be used. Furthermore, at the explore stage, the researcher describes the dataset of the sample data that has gone through data crawling. At the modify stage, the researcher will perform pre-processing, which is useful for selecting and cleaning raw data from several parts that can interfere with the process and changing variables to focus the model selection process through data cleaning, tokenizing, normalizing, stopword removal, and stemming. Then, there is a model stage where the researcher weights words using TF-IDF, then models the data by labeling data using a lexicon dictionary and classifying the data using the Naïve Bayes, K-Nearest Neighbor, and Support Vector Machine algorithms. At the final stage, there is an assessment discussion regarding model evaluation using the method and confusion matrix with true positive, true negative, false positive, and false negative indicators.

The data collection methods that will be used in this study are scraping, literature study, and literature study. The research instruments used in this study are tools and literature. In this study, the research device that will be used is hardware and software as tools and materials needed to facilitate the implementation of this study. This study uses a research methodology that refers to the SEMMA Data Mining Process method. The SEMMA method consists of sample, explore, modify, model, and assess.

3 Result and Discussion

3.1 Result

3.1.1 Sample

The first step is selecting a sample of data. The goal here is to choose a representative sample from the overall dataset. Sampling is necessary to ensure that the dataset used in the analysis is not too large, saving time and computational resources. However, the sample must be representative enough to reflect the patterns present in the entire dataset.

SAGA: Journal of Technology and Information Systems
Vol 2, Issue 3, August 2024, Page 288-294
ISSN: 2985-8933 (Media Online)
DOI: 10.58905/SAGA.v2i3.334

3.1.1.1 Crawling Data

a. Access Twitter's Application Programming Interface (API) to retrieve datasets directly from the platform.
b. Install Snscrape and Pandas, then import dependencies into the Google Collaboration notebook.
c. Through crawling data according to the search data being searched for. Then the results of the crawling data frame are in the form of a database that is stored in a Comma-Separated Values file, better known as CSV.

3.1.2 Explore

In this stage, data is explored using statistical techniques and visualization to identify patterns, anomalies, or relationships among variables. Data exploration is essential to gain initial insights into the structure and characteristics of the data[11]. The crawling data results contain 3 columns consisting of tweets, which are tweets from Twitter users; usernames, which contain the Twitter user account name; and created_at, which is the user's time when posting the tweet. The crawling results produced 4,347 tweets, which were used as the analysis material. The data was taken from Twitter users who tweeted their opinions or sentiments about the Erspo jersey, either in positive, negative, or neutral forms.

3.1.3 Modify

The modify step involves preparing the data for modeling. This includes cleaning the data, handling missing values, transforming variables, removing duplicates, and feature engineering. This stage ensures that the data is of high quality and relevant to the analysis objectives[12].

3.1.3.1 Cleaning Data

Data cleaning is the process of cleaning data from tweets that have been obtained. Before the cleaning process, the raw data from the crawling results still contains many elements that can affect the analysis results. After the cleaning process is carried out, the clean tweet results will be more ready to be analyzed. At this stage, tweets that have gone through the cleaning process will only contain core text that reflects user opinions.

3.1.3.2 Tokenizing

Tokenizing is the process of breaking down tweets into smaller units called "tokens." Before the tokenizing process, tweets are usually still in their entirety in the form of sentences or paragraphs. After the tokenizing process is carried out, tweets are broken down into separate words or units. Thus, each word or token can be explained individually to identify the meaning and sentiment behind it.

3.1.3.3 Normalize

Normalize or data normalization is a process in data processing to change data values into a certain scale or range so that the data becomes more consistent, uniform, and ready to be analyzed or used in the model. In words that have been tokenized, we may find various forms of words or informal terms used by Twitter users. For example, the word "yg" is an abbreviation of "yang" or informal words such as "keknya," which actually refers to the word "kayaknya." This kind of variation is often a challenge in text analysis because each variation can be understood as a different word, even though it has the same meaning. Therefore, normalization is carried out to change these variations into a more common and easily recognizable standard form.

3.1.3.4 Stopword Removal

Stopword removal is the process of removing common or frequently occurring words in a text that do not provide much meaning or important information value. This process makes the sentence more concise, leaving only the important words that reflect the sentiment of the tweet. The stopword removal process also helps in reducing noise in the data so that the algorithm can focus more on relevant patterns and keywords. The end result of this process is text that is more efficient and easier for machine learning algorithms to process, which is more sensitive to words that contain important information.

SAGA: Journal of Technology and Information Systems
Vol 2, Issue 3, August 2024, Page 288-294
ISSN: 2985-8933 (Media Online)
DOI: 10.58905/SAGA.v2i3.334

3.1.3.5 Stemming

Stemming aims to change words into their basic form (root words) by removing affixes such as prefixes, suffixes, or infixes. The results of the stemming process show the difference between tweets before and after data cleaning. In raw data, tweets usually contain words that are not very important for analysis, such as "yang," "dan," and less relevant word forms. After cleaning, tweets become more concise and focus on the core message.

3.1.4 Model

At this stage, modeling algorithms are applied to the modified data. The goal of modeling is to make predictions or classifications based on the patterns found in the data. Various machine learning algorithms, such as Naïve Bayes, Support Vector Machine (SVM), and K-Nearest Neighbor (KNN), are commonly used in this step[13]. This stage is passed by weighting words using TF-IDF. Furthermore, data classification is carried out using the lexicon-based method based on the lexicon dictionary that has positive and negative values. Tweet data that has been labeled with sentiment is then classified using machine learning algorithms, namely the Naïve Bayes Classifier, K-Nearest Neighbor, and Support Vector Machine methods.

3.1.5 Asses

The final stage in the SEMMA method is the assessment of the model. The results of the model are analyzed to evaluate how well it predicts or classifies the data. Common evaluation metrics include accuracy, precision, recall, and area under the ROC curve (AUC)[14].

3.1.5.1 Naïve Bayes Classifier

Based on the calculation results of the confusion matrix, this model has varying performance in detecting positive and negative classes. The model accuracy of 68.89% indicates that the model is able to provide correct predictions on around 68.89% of the total data tested. Although the accuracy is quite good, there is an imbalance in the model's performance between the positive and negative classes.

For the positive class, the model has very good performance. The positive precision is 76.51%, meaning that of all predictions classified as positive, around 76.51% of them are truly positive. The recall for the positive class is very high, reaching 100%, indicating that the model successfully detects all positive cases without missing any (false negatives = 0). This is reflected in the positive class f1-score value of 86.69%, indicating a very good balance between precision and recall in detecting the positive class.

However, for the negative class, the model's performance is very low. Although the negative precision reaches 100%, meaning that every time the model predicts negative, the prediction is always correct, the recall for the negative class is only 10.25%, indicating that the model fails to detect most of the negative cases, so that many negative cases are wrongly classified as positive (false positives are very high). This is also evident from the very low f1-score of the negative class, which is 22.60%, indicating that the model is not effective in predicting the negative class consistently. This performance imbalance, particularly between precision and recall for different classes, is a common issue in binary classification, especially when dealing with imbalanced datasets. In such cases, metrics such as precision, recall, and F1-score provide more insight into model performance than accuracy alone, as accuracy can be misleading in cases where one class dominates the other[15].

3.1.5.2 K-Nearest Neighbor

Based on the calculation results of the confusion matrix, this model shows quite good performance but with some weaknesses that need to be considered. The model accuracy of 63.66% shows that the model is able to make correct predictions on around 63.66% of the total data, indicating that the model can still be improved, especially in terms of predicting certain cases.

For the positive class, the model has a precision of 82.40%, which means that of all predictions classified as positive, around 82.40% are truly positive. The recall for the positive class is very high, at 95.69%, indicating that the model is able to detect almost all positive cases with very few false negatives. This is also reflected in the positive class f1-score value of 88.53%, which shows a good balance between precision and recall in predicting the positive class.

SAGA: Journal of Technology and Information Systems
Vol 2, Issue 3, August 2024, Page 288-294
ISSN: 2985-8933 (Media Online)
DOI: 10.58905/SAGA.v2i3.334

However, for the negative class, the model's performance is less than optimal. Although the precision for the negative class is quite good at 83.33%, meaning that most of the negative predictions are correct, the recall for the negative class is much lower at 51.28%. This shows that the model fails to detect more than half of the true negative cases (there are many false positives), so some negative cases are incorrectly classified as positive. The F1-score for the negative class is also relatively low at 63.48%, indicating that the model is not very effective in consistently predicting the negative class. The imbalance between precision and recall, especially in the negative class, suggests the need for adjustments in the model's threshold or the consideration of different algorithms or techniques to better handle class imbalances. A study discusses the relationship between precision-recall and ROC curves, explaining how varying the threshold can help balance these metrics depending on the specific needs of the application. This approach allows for a more tailored optimization of performance, particularly in cases of imbalanced data where certain metrics may need to be prioritized[16].

3.1.5.3 Support Vector Machine

Based on the above metrics calculated from the confusion matrix, this model has excellent performance in predicting both positive and negative classes. The model accuracy of 96.15% indicates that the model is able to provide correct predictions on most of the data tested.

For the positive class, the precision reaches 96.65%, which means that most of the positive predictions made by the model are correct. In addition, the recall of the positive class is 100%, indicating that the model does not miss a single positive case (no false negative or FN).

For the negative class, the model also shows perfect precision, which is 100%, meaning that there are no wrong negative predictions (no false positive or FP). The recall for the negative class is 91.60%, which means that there are some negative cases that are misclassified as positive, but most negative cases are successfully detected correctly.

The f1-score value for the positive class of 98.29% reflects an excellent balance between precision and recall, indicating that the model works very effectively in detecting the positive class with very few errors. Likewise, the f1-score for the negative class of 95.61% shows excellent performance in detecting the negative class despite slightly lower recall. Discuss common issues in classification with imbalanced datasets, where one class is more dominant than the other. The study highlights that in the presence of class imbalance, accuracy can often be a misleading metric. The researchers recommend using metrics like precision, recall, and F1-score to provide a more realistic evaluation of model performance. In your case, the high F1-scores for both positive and negative classes indicate a good balance between precision and recall, despite the slightly lower recall for the negative class[17].

3.2 Discussion

3.2.1 Social Media User Sentiment X On Erspo Jersey

The analysis in this study uses the Jupyter Notebook tool with the Python programming language. The stages carried out in this study include crawling tweet data, text pre-processing, labeling datasets using lexicon dictionaries, data classification using 3 classification methods, namely Naïve Bayes, K-Nearest Neighbor, and Support Vector Machine, testing using a confusion matrix, and visualization using Wordcloud.

The data retrieval process produces a csv file containing 15 detailed columns, namely conversation_id_str, created_at, favorite_count, full_text, id_str, image_url, in_reply_to_screen_name, lang, location, quote_count, reply_count, retweet_count, tweet_url, user_id_str, and username. Researchers in this study only used 3 columns out of 15 columns obtained, namely full_text, username, and created_at. Then the tweets are sorted from the newest to the earliest. The time span is from January 1, 2024, to August 31, 2024. There is a visualization of 5 of the 4,347 tweet data obtained from that time span. Furthermore, the data that has been obtained from crawling the tweet data is subjected to a data preprocessing stage. The dataset that has gone through several stages of data pre-processing is then called clean tweet data, which finally amounts to 1,717 tweet data.

The data that has been cleaned and structured by carrying out the TF-IDF weighting process. Furthermore, the labeling process is carried out, namely labeling the data into 3 sentiment classes, namely positive class, negative class, and neutral class. Class labeling in this dataset uses one of the modeling stages, namely lexicon-based. The tweet data will be compared and given a value or score according to the lexicon dictionary. The sum of these scores will determine whether the tweet data has a positive class label and a negative class.

By using the Python programming language, researchers can show wordcloud results in the form of words that are often spoken in tweets. This is supported by research that emphasizes sentiment as a reflection of a person's attitude or feelings towards something, whether positive, negative, or neutral. Understanding sentiment

SAGA: Journal of Technology and Information Systems
Vol 2, Issue 3, August 2024, Page 288-294
ISSN: 2985-8933 (Media Online)
DOI: 10.58905/SAGA.v2i3.334

helps reveal the emotions underlying opinions, offering a clearer perspective on how people respond to specific topics[18].

3.2.2 Performance of Sentiment Results in the Application of Machine Learning Algorithms

In this study, the last stage is the assessment stage. This stage evaluates 3 classification models of machine learning algorithms, namely Naïve Bayes, K-Nearest Neighbor, and Support Vector Machine. The evaluation contains the confusion matrix value, accuracy value, precision value, recall value, and f1-score value of the test data and training data on each of the 3 classification models of the machine learning algorithm used. The following is a summary of the comparison of the results of the accuracy, precision, recall, and f1-score values ofhe 3 classification methods used in this study.

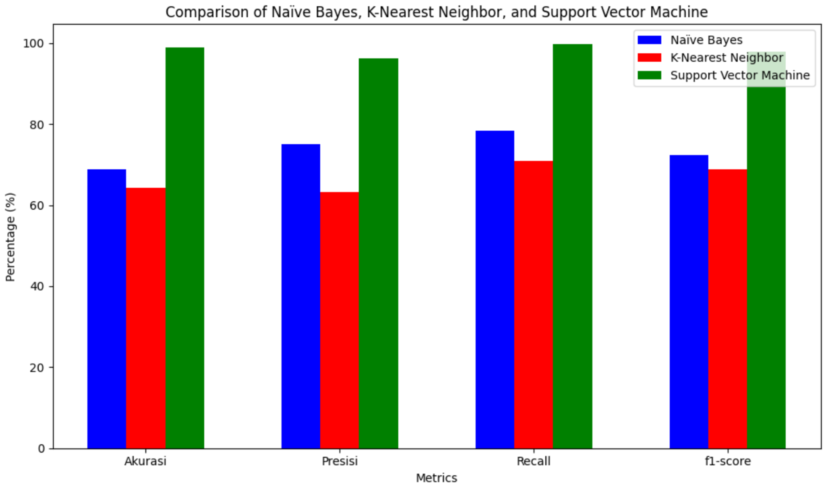

Figure 1. Comparison of Accuracy, Precision, Recall, and f1-score Values

In the image above, you can see each accuracy value, precision value, recall value, and f1-score value from the ratio previously determined by the researcher. The image above shows a comparison of the performance of three machine learning algorithms, namely Naïve Bayes, K-Nearest Neighbor (KNN), and Support Vector Machine (SVM).

From the image, it can be seen that Support Vector Machine (SVM) consistently shows the best performance in all metrics. With an accuracy approaching 99%, SVM far surpasses Naïve Bayes and KNN. The precision and recall of SVM are also very high, above 96% and 99%, respectively, indicating a strong ability to identify positive classes and minimize prediction errors. Naïve Bayes shows better performance than KNN in precision, recall, and f1-score metrics, although it still lags behind SVM. Naïve Bayes has an accuracy of around 68.89%, with a precision of 74.98% and a recall of 78.28%. K-Nearest Neighbor (KNN) has the lowest performance among the three models, with an accuracy of 64.19%. This shows that KNN makes more mistakes in classifying data than the other two algorithms. KNN's precision is also low, only around 63.15%, indicating that there are many false positive predictions.

Overall, Support Vector Machine (SVM) is clearly the most effective model in predicting with minimal error rates, while K-Nearest Neighbor tends to give lower results than the other two algorithms.

4 Conclusion

Based on the results of the study, it can be concluded that positive sentiment dominates tweets about Jersey Erspo, especially related to the Indonesian national team jersey product. Positive sentiment dominates tweets about Jersey Erspo, especially related to the Indonesian national team jersey product. Then, of the three algorithms tested, Support Vector Machine (SVM) gave the best performance results with an accuracy of 98.97%, a precision of 96.27%, a recall of 99.67%, and an f1-score of 97.85%. Meanwhile, K-Nearest Neighbor (KNN) with the best k value (k = 10) had an accuracy of 64.19%, and Naïve Bayes produced an accuracy of 68.89%. SVM proved to be the most superior classification method in this study compared to KNN and Naïve Bayes.

5 Acknowledgementds

The researcher would like to express his gratitude to his family, supervisor, Platform X, and colleagues who have provided support, guidance, and valuable input during the process of compiling this research.

References

[1] We Are Social 2023 Social Media Marketing Benchmark Report 2024 Influencer Marketing Hub Diakses pada September 2023 dari https://influencermarketinghub.com/social-media-marketing-benchmark-report/
[2] Putri D, Nama G F, Sulistiono W E 2022 Analisis Sentimen Kinerja Dewan Perwakilan Rakyat (DPR) Pada Twitter Menggunakan Metode Naive Bayes Classifier Jurnal Informatika Dan Teknik Elektro Terapan **10** 34-40 https://doi.org/10.23960/jitet.v10i1.2262
[3] Verawati I, Audit B S 2022 Algoritma Naïve Bayes Classifier Untuk Analisis Sentimen Pengguna Twitter Terhadap Provider By.U Jurnal Media Informatika Budidarma **6** 1411 https://doi.org/10.30865/mib.v6i3.4132
[4] Zhang Chen, Liu Xi 2020 A Distributed Storage And Computation K-Nearest Neighbor Algorithm Based Cloud-Edge Computing For Cyber-Physical-Social-Systems IEEE Access https://doi.org/10.1109/Access.2020.297476
[5] Kumar R et al. 2023 Sentiment Analysis of Government Policies Using SEMMA Method Journal ABC
[6] Brahimi B, Touahria M, Tari A 2021 Improving sentiment analysis in Arabic: A combined approach Journal of King Saud University - Computer and Information Sciences **33** 1242-1250 https://doi.org/10.1016/j.jksuci.2019.07.011
[7] Wicaksono N B A 2022 Analisis sentimen tingkat kepuasan pengguna BPJS dengan metode SentiStrength Jurnal Ilmiah Intech: Information Technology Journal of UMUS **4** 189-199 https://doi.org/10.46772/intech.v4i02.868
[8] Irnawati O, Solecha K 2022 Analisis sentimen ulasan aplikasi Flip menggunakan Naïve Bayes dengan seleksi fitur PSO Jurnal Ilmiah Intech: Information Technology Journal of UMUS **4** 189-199 https://doi.org/10.46772/intech.v4i02.868
[9] Moleong L J 2016 Metodologi penelitian kualitatif (Revised ed.) PT Remaja Rosdakarya
[10] Arikunto S 2014 Prosedur penelitian: Suatu pendekatan praktik (Revised ed.) Rineka Cipta
[11] Larose D T, Larose C D 2015 Data Mining and Predictive Analytics John Wiley & Sons
[12] Han J, Kamber M, Pei J 2011 Data Mining: Concepts and Techniques Morgan Kaufmann
[13] Hastie T, Tibshirani R, Friedman J 2009 The Elements of Statistical Learning: Data Mining, Inference, and Prediction Springer
[14] Witten I H, Frank E, Hall M A 2011 Data Mining: Practical Machine Learning Tools and Techniques Morgan Kaufmann
[15] Saito T, Rehmsmeier M 2015 The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets PLOS ONE **10** e0118432 https://doi.org/10.1371/journal.pone.0118432
[16] Davis J, Goadrich M 2006 The relationship between Precision-Recall and ROC curves Proceedings of the 23rd international conference on Machine learning 233-240
[17] He H, Garcia E A 2009 Learning from imbalanced data IEEE Transactions on Knowledge and Data Engineering **21** 1263-1284
[18] Liu B 2015 Sentiment analysis: Mining opinions, sentiments, and emotions Cambridge University Press http://ejournal.uajy.ac.id/27165/http://ejournal.uajy.ac.id/27165/3/150708283_Bab2.pdf