Machine learning approach to customer sentiment analysis in twitter airline reviews

Ekka Pujo Ariesanto Akhmad[1]*, *Kusworo* Adi[2], and *Aris* Puji Widodo[3]

[1]Doctoral of Information System Department, Diponegoro University, Semarang, Indonesia
[2]Physics Department, Science and Math Faculty, Diponegoro University, Semarang, Indonesia
[3]Informatics Department, Science and Math Faculty, Diponegoro University, Semarang, Indonesia

Abstract. Customers typically provide both online and physical services they use ratings and reviews. However, the volume of reviews might grow very quickly. The power of machine learning to recognize this kind of data is astounding. Numerous algorithms that could be employed for job of sentiment analysis have been developed to categorize tweets about airline sentiment into positive, neutral, or negative categories, this study compares the effectiveness algorithm for machine learning Naive Bayes (NB), Logistic Regression (LR), Decision Tree (DT), Support Vector Machine (SVM), Adaboost, Extreme Gradient Boosting (XGB), Light Gradient Boosting Machine (LGBM), and Random Forest (RF) dividing the Twitter airline sentiment data into positive, neutral, or negative categories using the TF IDF model. The experiment involved two phases of activity: a classification algorithm utilizing SMOTE and sans SMOTE with Stratified K-Fold CV algorithm. With the RF model, the greatest performance accuracy for SMOTE is 97.56%. Without SMOTE, the RF with a value of 92.21% provides the maximum performance accuracy. The findings demonstrate that SMOTE oversampling can improve sentiment analysis accuracy.

Keywords. airline reviews, sentiment analysis, machine learning, SMOTE, stratified k-fold CV

1. Introduction

High-speed Internet access is now widely available, which has changed how people choose which airlines to fly with and how information about travel services is disseminated [1,2]. Travel planning is dominated by online reviews since they offer trustworthy information about destinations before a trip [3].

Numerous firms have paid close attention to social media data, order data, and sentiment analysis using online reviews [4]. Online reviews of social media mining are numerous; connectivity, content, and user data are a few examples [5]. A notable contribution to the field of travel and tourism management is the evaluation and qualitative analysis of social media information content to extract key components [6]. Therefore, it is vital for the travel sector to assess and forecast customer suggestions utilizing online reviews [7].

Online reviews containing user-generated content have increasingly gained popularity, posing a fresh management and monitoring issue for the business. It will be challenging for the buyer to choose wisely under these situations. To extract helpful patterns, many service providers employ a range of techniques. Customers publish reviews and rate specific services like seat comfort, the caliber of food and beverages, and other characteristics based on the business itself on online review and rating sites. While the rankings are presented

* Corresponding author: eka.pujo@hangtuah.ac.id

numerically, the reviews are provided in literary form. In most cases, the textual data, rather than the numerical ratings, contains the descriptive and emotive information about the whole service. However, linguistic data has a higher potential for human error than straightforward value ratings from 1 to 5. Furthermore, non-English speaking nations are more prone to making blunders of this nature. Because of this, the majority of studies emphasized quantitative rather than qualitative content. [7,8].

In order to extract insights from qualitative data, NLP, and Text Mining (TM), many computer models are utilized in sentiment analysis [9]. The approach for sentiment analysis commonly makes use of machine learning algorithms, including Deep Learning, unsupervised learning, and ensemble learning, as well as lexicon-based and hybrid methods. [10].

A framework for conducting sentiment analysis on airline databases is suggested in this paper. KAGGLE database information was acquired [11]. There are several machine learning (ML) techniques employed, including LR, NB, SVM, DT, Ensemble Learning Boosting (Adaboost, LGBM, XGB), and RF. Also contrasted are the outcomes. Our primary contribution is to use SMOTE oversampling and Stratified KFold Cross Validation to compare the performance of the RF framework to that of current ML techniques.

The remainder of the piece is organized as follows. The relationship between earlier research on Customer Sentiment Analysis of Airline Reviews is explained in Section 2. Additionally, Section 3 covers the technique. In Section 4, there is a description of experiments and test findings. We sum up our findings and future research initiatives in Section 5.

2. Related Work

2.1 Machine learning for airline reviews

For sentiment analysis and user recommendations, many researchers have used machine learning (ML) techniques.

[1] provides a clear explanation of how online reviews can be used to measure and forecast consumer mood. To distinguish between the core and additional services in the study's qualitative contents, aspect level sentiment analysis was used. To bolster their findings, the authors also compared ML models like neural networks (NN), NB, and SVM. Their findings demonstrated how accurately their NN-based model predicted customer recommendations.

The sentiment in the airline tweet dataset was examined by the authors of [12] using aspect-based sentiment analysis. The authors' model for aspect detection and polarity identification utilized the SVM technique. It was demonstrated that performance increased when word-embedding was employed for feature extraction on a dataset of airline tweets, on which the SVM model was trained and tested.

The dataset was gathered from six US airline companies, and many machine learning (ML) techniques, including RF, DT, LR, K-nearest Neighbor (KNN), SVM, and AdaBoost, were trained and tested on it in the study. [13]. In their implementation, the training set comprised 80% of the data, while the testing set comprised the remaining 20%. The findings demonstrate that SVM and LR successfully classified sentiment into three classes with an accuracy of more than 80%.

The analysis of the tweets in this study uses machine learning to enhance the user experience. Utilizing word embedding, the Glove dictionary approach, and the n-gram approach, features were retrieved from the tweets. In order to create a classification model that divides tweets into positive and negative categories, SVM and a number of ANN (artificial neural network) designs were also taken into consideration. Convolutional neural networks (CNNs) were also created to classify the tweets, and the outcomes were compared to the most precise models created using SVM and other ANN designs. It was discovered that CNN performed better than SVM and ANN models. To map the relationship with emotion categories, association rule mining has been done on a variety of tweet kinds. The findings reveal a number of significant connections that undoubtedly aid airline companies in enhancing the customer experience [14].

In the fields of machine learning (ML) and natural language processing (NLP), sentiment analysis has gained popularity. Deep Learning (DL) methods are now used to obtain accurate sentiment analysis results. In this study, a hybrid CNN-LSTM (long short-term memory-convolutional neural network) model for sentiment analysis was presented. To achieve the required outcomes, the proposed model is implemented using batch normalization, dropout, and max pooling. Datasets from Twitter and Airlinequality were used in an experimental analysis of airline sentiment. Utilizing close spacing between related words, the Keras word embedding approach was used to turn texts into vectors of numerical values. To evaluate the effectiveness of the model, they computed a number of factors, including accuracy, precision, recall, and F1-measure. These model parameters outperform traditional ML models for sentiment analysis. The suggested model excels in sentiment analysis with 91.3% accuracy, according to their study of the results. [15].

2.1.1 Ensemble learning boosting

This study uses a supervised machine learning method, namely ensemble learning boosting [16], which states that this method works by boosting a weak initial classification model. Model strengthening is performed sequentially using bootstrap data object sampling based on dynamic weighting. The process of strengthening the classification model sequentially is carried out T times until a classification model is produced that is considered strong enough. A number of T models produced are then combined using a majority vote with weighting according to accuracy, not one man one vote as in Bagging. In this Boosting method, the model that has the highest accuracy gets the largest weight while the model that has the lowest accuracy gets the smallest weight.

2.1.2 Random Forest

The Random Forest (RF) method is a variant of Bagging [17]. RF is a combination of decision trees such that each tree depends on independently sampled random vector values with the same distribution. RF uses random feature selection to sort out each vertex so as to produce relatively high accuracy. The difference between RF and Bagging lies in the number of attributes used. Bagging uses all the attributes to build an independent model while RF only uses some of the features. Thus RF is more efficient in computing. The set of independent models built by RF is also more varied compared to Bagging.

2.2 SMOTE (Synthetic Minority Over-sampling Technique)

One of the most popular over-sampling methods, SMOTE generates synthetic data in minority data classes to accommodate unbalanced datasets. This will balance the data [18] which will help with better classification performance [19] among other benefits.

The following is the formula for creating synthetic data using SMOTE.

$$X_{new} = X_i + (\hat{X}_k - X_i) \text{ x } \delta \quad (1)$$

where X_{new} = new synthetic data, X_i = data from the minority class, $\hat{X}_k$ = data from k nearest neighbor that has the closest distance to X_i, and δ = random number between 0 and 1. The difference in distance in determining the nearest neighbor in numerical data is done by using Euclidean distance.

2.3 Stratified k-Fold Cross-Validation

Stratified K-Fold Cross-Validation (SKCV) extends Cross-Validation (CV). Each class is distributed equally across the k-fold thanks to the SKCV, claims [20]. To put it another way, the datasets are split into k-folds without affecting the SKC's sample distribution ratio for any class. By using stratified sampling as opposed to random sampling, SKCV was recommended by [21] to guarantee that relative class frequencies are successfully maintained over each train and validation cycle. The cervical cancer data set is divided into k groups or folds of roughly equal size using a stratified sampling approach in this procedure. Therefore, when dealing with classification issues involving unequal class distributions, SKCV is favored over CV. The sample distribution ratio inside SKCV for all classes is shown in Fig. 1.

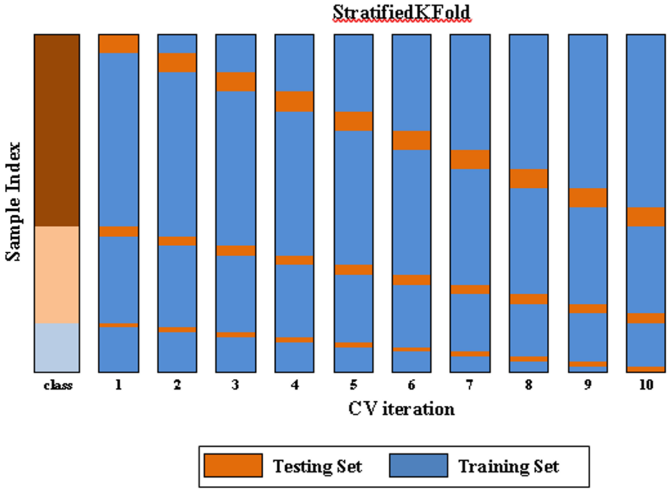

Fig. 1. The sample distribution ratio inside SKCV

3. Research Method

The data collecting, feature engineering, preprocessing, feature extraction, classification model building, and performance evaluation are the six fundamental activities that make up this study. A variety of standard machine learning methods are used to create classification models. Fig. 2 depicts the research approach employed in this study.

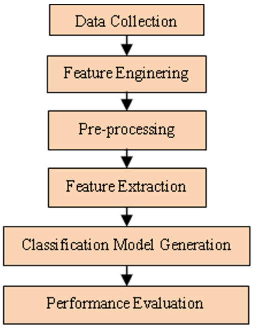

Fig. 2. Research Diagram Block

3.1 Data Collection

For this study, airline reviews were acquired from Kaggle [11]. For the sentiment classification task on each large U.S. airline, we simply use the reviews and the sentiment label from the dataset for all rows. The data provided comes from passengers or travelers who have expressed their feelings or thoughts regarding the airlines they have used. The data is divided into categories such airline feelings (good, negative, or neutral), a list of airlines that passengers have flown on, traveler and passenger characteristics, location, the time at which the comment was written, the time zone, and the passenger's comment or text [22]. Data in this dataset will be interpreted negatively in 63% of the cases, neutrally in 21% of the cases, and positively in 16% of the cases. The distribution of the data can be used to infer if the dataset is unbalanced.

In this investigation, there was an imbalance in the number of tweets for each label in the data distribution. The performance of classification may suffer if the dataset is uneven [23]. To do that, we use the SMOTE (Synthetic Minority Oversampling Technique) to alter a dataset and overcome the issues with balance data.

3.2 Feature Engineering

Perform feature engineering consisting of categorization null values, checking duplicate in review text, and calculates the total tweet of each sentiment.

3.3 Preprocessing

Process the text for the initial evaluation. The accuracy of the system is improved by the use of text processing, which helps to better arrange the data. Tokenization, filtering, stemming, and case folding are a few of the sub-processes. After that, the data is prepared for weighing. [24].

3.3.1 Case Folding

The case folding method involves changing each word in a phrase to lowercase, adding a space, and a full stop. This is crucial since it frequently occurs for a tweet to begin with a capital letter in its entirety or for a typing error to cause a capital letter to appear in the middle of a word. By doing this, the system won't be case-sensitive.

3.3.2 Tokenizing

Tokenizing is a technique that converts a statement into a group of words.

3.3.3 Filtering

Filtering is a technique that gets rid of words that aren't as important or, if they don't exist, don't alter the meaning of the text.

3.3.4 Stemming

Affix-containing words will now be changed into the base word. As a result, words that have the same meaning cannot be misunderstood as having a distinct meaning just because of their affix.

3.4 Feature Extraction

The two most popular methods for numerically representing a text are Term Frequency (TF), which indicates how frequently a word appears in a tweet relative to how frequently it appears overall in the dataset,

$$TF = \frac{\text{Number of times the term appears in the document}}{\text{Total number of terms in the document}} \tag{2}$$

and Inverse Document Frequency (IDF), which indicates how significant a word is overall in the dataset [25].

$$IDF = log\left(\frac{\text{number of the documents in the corpus}}{\text{number of documents in the corpus contain the term}} \right) \tag{3}$$

At the word weighting stage, the Term Frequency –Inverse Document Frequency (TF-IDF) method is used to get the weight value for each word in the data used. The word weighting process uses the TF-IDF algorithm. TF-IDF presents a word frequency score especially for interesting words, for example words that appear frequently in one document but not for all documents. This process is carried out by calculating the weight of each word in the training data using the sklearn library. The TF-IDF of a term is calculated by multiplying TF and IDF scores.

$$TF\text{-}IDF = TF * IDF \tag{4}$$

3.5 Classification model generation

Several traditional machine learning and ensemble learning boosting techniques were used in this investigation. Used were the LR, NB, SVM, DT, Adaboost, LGBM, XGB, and RF classical machine learning algorithms [25].

The learning process takes input in the form of a collection of labeled training data (with class properties) and outputs a classification model [16].

3.6 Evaluation

The data was split into training and testing using a value of 10 for stratified K-fold cross-validation. Biased performance metric values can be avoided via stratified 10-fold cross-validation [18]. Training data and testing data are separated from the data. To create a classification model, which is then tested using test data, the training procedure is carried out using the training data. The fold value is increased through this procedure until it reaches 10.

A confusion matrix is used to assess the effectiveness of the classification model using metrics including accuracy, precision, recall, specificity, and F1 score [26]. The confusion matrix's structure is displayed in Table 1.

Table. 1. Confusion Matrix

Predicted	**Actual**	
	Positive	**Negative**
Positive	TP	FP
Negative	FN	TN

TP, True Positive; FN, False Negative;
FP, False Positive; TN, True Negative

This study uses accuracy to measure how accurately the model can correctly classify data.

$$\text{Accuracy} = \frac{\text{TP + TN}}{\text{TP + FP + FN + TN}} \tag{5}$$

Precision describes the level of accuracy between the requested true positive prediction data and the predicted results given by the model.

$$\text{Precision} = \frac{\text{TP}}{\text{TP + FP}} \tag{6}$$

Recall describes the success of the model in retrieving information.

$$\text{Recall} = \frac{\text{TP}}{\text{TP + FN}} \tag{7}$$

F1 score is a comparison of the weighted average precision and recall.

$$\text{F1 score} = 2 \text{ x } \frac{\text{Recall x Precision}}{} \tag{8}$$

Recall + Precision

4. Result and Discussion

The findings of the experiment and analysis are covered in this section.

4.1 Research Data

Data from 14,640 airline reviews was used in this study. The distribution of the data's specifics are as follows: 2,363 reviews of the data are rated positively, 3,099 are rated neutrally, and 9,178 are rated negatively. Analyze the information you found on the Kaggle website. Python 3 was used to carry out this study, together with the TF-IDF, SMOTE, LR, NB, SVM, DT, Adaboost, LGBM, XGB, and RF libraries for classifier model development. The effectiveness of the model was evaluated using stratified K-fold cross-validation. Ten is the used number of K. Then, 10 data sets are created, each of which has 8 folds of training data and 1 fold each of validation data and test data.

4.2 Research Scheme

We divided the experiment into two tests, one using SMOTE oversampling and the other using the ML Classifier and the Stratified K-Fold CV method without SMOTE oversampling. The TF-IDF technique was used to create a model that forecasts sentiment as well.

Table 2 displays the outcomes of several analyses of machine learning algorithms employing the TF-IDF algorithm feature with weighted using f1-weighted. The RF algorithm achieved accuracy values of 97.56% with SMOTE oversampling and 92.21% without SMOTE oversampling, respectively, for the best results. As a result, the experimental outcomes with the RF method have proved valid for identifying sentiment thus far. The measurement results are LGBM with an accuracy value of 93.51% when the Stratified K-Fold CV method is applied to the Classifier on data without SMOTE oversampling.

4.3 Discussion and Limitation

The RF algorithm has been successful in appropriately classifying attitudes that have been discovered during the data preparation stage based on the experimental findings presented in the preceding section. Using an accuracy value of 97.56% as opposed to the XGB and LGBM algorithms' respective accuracy values of 94.51% and 94.05% In contrast, the accuracy achieved for the procedure without SMOTE is also rather high, coming in at 93.51% as opposed to SMOTE's accuracy of 97.56% (see Table 2).

When viewed from the accuracy and f1-score, the ability of the LGBM method is better in classifying sentiment on imbalanced data or not using SMOTE compared to other ML methods. If the classifier model uses SMOTE, then SMOTE is able to improve the accuracy of minority class classification and avoid overfitting. Because the risk of overfitting is lower, the accuracy of the classifier model is also better.

Data imbalance is the fundamental problem with this approach, especially for class neutrals. Despite this, both the negative and positive classes in our model produce good outcomes (see Table 3). This study's focus is on predicting the explicit feelings of social media messages; however, it is not intended to anticipate the implicit sentiments found in texts that are sarcastic.

5. Conclusion

Based on the outcomes of our tests, we find that adding SMOTE oversampling and stratified k-fold cross validation yields a 97.56% accuracy value for the RF algorithm. When using the same algorithm, RF, the research sans SMOTE had an accuracy rate of 92.21%.

Table 2 shows that the LGBM sentiment classifier outperforms the other seven well-known classifiers for sentiment analysis (LR, SVM, RF, NB, Adaboost, XGB, and DT) in terms of accuracy and F1-score. By producing the best F1-score for the minority class, LGBM has shown that it is capable of handling the issue of class imbalance.

Table 2. Evaluation of TF-IDF feature performance on eight classifiers using F1-score and accuracy

ML Algorithms	Without SMOTE			With SMOTE		
	Accuracy	Macro Avg	Weighted Avg	Accuracy	Macro Avg	Weighted Avg
LR	92.2814	0.86	0.91	92.6262	0.91	0.91
NB	72.8142	0.52	0.65	88.4489	0.86	0.86
SVM	92.5546	0.86	0.92	95.0599	0.94	0.94
DT	91.3251	0.80	0.89	96.2586	0.95	0.95
Adaboost	90.5054	0.81	0.89	83.6905	0.82	0.82
LGBM	**93.5109**	**0.87**	**0.93**	94.0428	0.93	0.93
XGB	93.3060	0.86	0.92	94.5150	0.93	0.93
RF	92.2131	0.84	0.90	**97.5662**	**0.97**	**0.97**

Table 3. Precision, Recall, and F1-score utilizing a stratified 10-fold cv for positive, negative, and neutral classes

ML Algorithms	Positive			Negative			Neutral		
	Precision	Recall	F1-score	Precision	Recall	F1-score	Precision	Recall	F1-score
LR	0.93	0.81	0.87	0.97	0.99	0.98	0.83	0.92	0.88
NB	0.88	0.90	0.89	0.82	0.95	0.88	0.90	0.73	0.81
SVM	0.96	0.88	0.92	0.98	1.00	0.99	0.89	0.95	0.92
DT	0.93	0.94	0.93	0.99	0.98	0.98	0.93	0.93	0.93
Adaboost	0.82	0.65	0.72	0.96	0.97	0.97	0.71	0.85	0.78
LGBM	0.91	0.89	0.90	0.98	1.00	0.99	0.89	0.90	0.90
XGB	0.92	0.87	0.90	0.98	1.00	0.99	0.88	0.92	0.90
RF	0.95	0.96	**0.96**	0.98	0.99	**0.99**	0.96	0.95	**0.95**

As a result of this study's single-minded concentration on the sentiment analysis of online airline evaluations, numerous aspects of the travel and tourism sector can be addressed in the future for further elaborating consumer sentiment. In this study, we solely take into account data that was gathered from internet sources, specifically in the form of English sentences. The sentiment analysis does not take into account consumer reviews that are written in other languages; however, future research may include textual reviews that are offered in many languages. Finally, future research can integrate different classification or prediction methods into Lexicon-based or Deep Learning sentiment analysis models.

References

1. Siering, Michael. Amit V. Deokar, Christian Janze. "Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews." Decis. Support Syst. **107**: 52-63, (2018).
2. Marco, Julio Navio. Luis Manuel Ruiz-Gómez, Claudia Sevilla-Sevilla. Progress in information technology and tourism management: 30 years on and 20 years after the internet-Revisiting Buhalis & Law's landmark study about eTourism. Tour. Manag. Vol **69**, pp. 460–470, Dec (2018).
3. Ukpabi D, S Olaleye, E Mogaji, H Karjaluoto. Insights into online reviews of hotel service attributes: A cross-national study of selected countries in Africa. Inf. Technol. Tour. pp 243–256, (2018).
4. Akshi Kumar, Geetanjali Garg. Systematic literature review on context-based sentiment analysis in social multimedia. Multimed. Tools Appl. **79**, 15349–15380, (2019).
5. Imene Guellil, Kamel Boukhalfa. *Social big data mining: A survey focused on opinion mining and sentiments analysis*. In Proceedings of the 2015 12th International Symposium on Programming and Systems (ISPS'15). IEEE, Los Alamitos, CA, 1-10, (2015).

6. Chih-Fong Tsai, Kuanchin Chen, Ya-Han Hu, Wei-Kai Chen. Improving text summarization of online hotel reviews with review helpfulness and sentiment. Tour. Manag. **80**, 104122, (2020).
7. Jain, Praphula Kumar. Ephrem Admasu Yekun, Rajendra Pamula, Gautam Srivastava. "Consumer recommendation prediction in online reviews using Cuckoo optimized machine learning models". Comput. Electr. Eng. 95 107397, (2021).
8. Moro S, Rita P, Coelho J. Stripping customers' feedback on hotels through data mining: The case of las vegas strip. Tour. Manag. Perspect. **23**:41–52, (2017).
9. Mika V. Mäntylä, Daniel Graziotin, Miikka Kuutila. The evolution of sentiment analysis-A review of research topics, venues, and top cited papers. Comput. Sci. Rev. **27**, 16–32, (2018).
10. Ligthart, Alexander. C. Catal, B. Tekinerdogan. Systematic reviews in sentiment analysis: a tertiary study. Artif. Intell. Rev. **54**:4997–5053, (2021).
11. Rustam, Furqan. Imran Ashraf, Arif Mehmood, Saleem Ullah, Gyu Sang Choi.Tweets Classification on the Base of Sentiments for US Airline Companies. Entropy, 21, 1078, (2019).
12. Sternberg, F., Hedegaard Pedersen, K., Ryelund, N. K., Mukkamala, R. R., Vatrapu, R. "*Analysing Customer Engagement of Turkish Airlines Using Big Social Data*". 2018 IEEE International Congress on Big Data (Big Data Congress), (2018).
13. Rane A, Kumar A. "*Sentiment classification system of Twitter data for US airline service analysis*." In: IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC). 1, IEEE; p. 769–73, (2018).
14. Kumar, Sachin. Mikhail Zymbler. "A machine learning approach to analyze customer satisfaction from airline tweets". J. Big Data 6, 1 62, (2019).
15. Jain, Praphula Kumar. Vijayalakshmi Saravanan, Rajendra Pamula. "A Hybrid CNN-LSTM: A Deep Learning Approach for Consumer Sentiment Analysis Using Qualitative User-Generated Contents". ACM Trans. Asian Low-Resour. Lang. Inf. Process. 20, 5, Article 84, 15 pages, July (2021).
16. Tan, K.L., Lee, C.P., Lim, K.M. A Survey of Sentiment Analysis: Approaches, Datasets, and Future Research. Appl. Sci. (2023), 13, 4550.
17. Breiman, L.E.O. Random Forests. Mach. Learn. **45**, p.5-32, (2001).
18. N. Chawla, K. Bowyer, L. Hall, W. Kegelmeyer, "SMOTE: Synthetic Minority Over-Sampling Technique," J. Artif. Intell. Res., vol. **16**, p. 321–357, (2002).
19. J. Ah-Pine, E. P. S. Morales, "*A Study of Synthetic Oversampling for Twitter Imbalanced Sentiment Analysis* ", Proceedings of the Workshop on Interactions between Data Mining and Natural Language Processing, DMNLP, (2016).
20. Allen, J., Liu, H., Iqbal, S., Zheng, D., Stansby, G. Deep learning-based photoplethysmography classification for peripheral arterial disease detection: A proof-of-concept study. Physiol. Meas. **42**(5), (2021).
21. Prusty S, Patnaik S, Dash SK. SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer. Front. Nanotechnol. 4:972421, (2022).
22. Patel, Aksh. Parita Oza, Smita Agrawal. Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model. Procedia Comput. Sci. 218 2459–2467, (2023).
23. Kumar, Pradeep. Roheet Bhatnagar, Kuntal Gaur, Anurag Bhatnagar. *Classification of Imbalanced Data:Review of Methods and Applications*. IOP Conf. Series: Materials Science and Engineering 1099 012077, (2021).
24. Jain, Praphula Kumar. Rajendra Pamula, Gautam Srivastava. "A systematic literature review on machine learning applications for consumer sentiment analysis using online reviews". Comput. Sci. Rev. **41** 100413, (2021).

25. Alzamzami, Fatimah. M. Hoda, A. El Saddik. “Light Gradient Boosting Machine for General Sentiment Classification on Short Texts: A Comparative Evaluation”. IEEE Access. May, (2020).
26. Fatemeh Hemmatian, Mohammad Karim Sohrabi. A survey on classification techniques for opinion mining and sentiment analysis. Artif. Intell. Rev. **52**, 1495–1545, (2019).