

Sentiment Analysis of Impact of Technology on Employment from Text on Twitter

<https://doi.org/10.3991/ijim.v14i07.10600>

Shahzad Qaiser

Capital University of Science and Technology (CUST), Islamabad, Pakistan

Nooraini Yusoff ^(✉)

Universiti Malaysia Kelantan, Kelantan, Malaysia

nooraini.y@umk.edu.my

Farzana Kabir Ahmad, Ramsha Ali

Universiti Utara Malaysia, Kedah, Malaysia

Abstract—Various studies are in progress to analyze the content created by the users on social media due to its influence and the social ripple effect. The content created on social media has pieces of information and the user's sentiments about social issues. This study aims to analyze people's sentiments about the impact of technology on employment and advancements in technologies and build a machine learning classifier to classify the sentiments. People are getting nervous, depressed, and even doing suicides due to unemployment; hence, it is essential to explore this relatively new area of research. The study has two main objectives 1) to preprocess text collected from Twitter concerning the impact of technology on employment and analyze its sentiment, 2) to evaluate the performance of machine learning Naïve Bayes (NB) classifier on the text. To achieve this, a methodology is proposed that includes 1) data collection and preprocessing 2) analyze sentiment, 3) building machine learning classifier and 4) compare the performance of NB and support vector machine (SVM). NB and SVM achieved 87.18% and 82.05% accuracy, respectively. The study found that 65% of people hold negative sentiment regarding the impact of technology on employment and technological advancements; hence, people must acquire new skills to minimize the effect of structural unemployment.

Keywords—Sentiment Analysis, Unemployment, Technology, Machine Learning, Natural Language Processing

1 Introduction

Technology is taking over the world in terms of jobs in all disciplines. Today, many of the jobs that humans used to perform are being performed by technologies like artificial intelligence [1], [2]. While technology is helping humanity to get things done faster with higher accuracy in minimum time and cost as predicted by [3], it is also causing humans to lose their jobs at a way faster rate than the new jobs are being created

leading to structural unemployment in the society as highlighted in the MIT technology review by [4]. World Economic Forum (WEF) has reported that 50% of the companies are expecting a reduction in the workforce by 2022 due to automation. According to the report, 42% of the work in companies will be done by the machines by the end of 2022 compared to 29% in 2018 [5].

As per [6], the report published by Gallup, the longer the people are not employed, it is more likely that they report signs of poor psychological health. As per the report, one in every fifth American who has not been employed for one year or more, have or is being treated for depression. 11.1% of people who are unemployed from last two weeks or less and 19% of people who are unemployed from the last 52 weeks or more are reported to have or being treated for depression in the US. According to studies [7] and [8], researchers who studied the relationship between suicides and unemployment are of the view that on average, unemployment is the cause of 45,000 suicides per year worldwide and the numbers are still increasing. Unemployment is one of the leading cause of suicides in the US, and people are concerned about suggesting measures to overcome the issue so that the number can be minimized [7], [8].

If people have negative sentiments regarding the technology advancements, it is safe to assume that such people are generally nervous, depressed, or concerned about losing their jobs to technology in the near future. If people are nervous or depressed, that means their psychological health is upset or going to be affected shortly once their employment status will be affected by technology, as seen in [6]. Similarly, if they are depressed, they may go toward attempting suicide due to long-time unemployment soon as it is observed by [8]. So, it is essential to save people from nervousness, depression, and suicide attempts. It is an urgent need to understand their sentiments regarding rapid developments in technology and how it is impacting their employment so that preventive measures can be taken. Once it is identified that people have negative sentiments regarding the technological developments, they can be guided to upgrade their skill set to remain relevant in this digital economy and the era of automation so that structural unemployment and its related concerns can be minimized. It is relatively a new area to explore from a technological point of view. There is a gap in this area as there is rarely any research that has performed sentiment analysis on the impact of technology on employment. However, a few studies, such as [1] has proposed a theoretical model to assess how artificial intelligence is shaping the job market and replacing humans. There can be many factors contributing to nervousness, depression, and suicides due to structural unemployment, but this study is focused on the technological impact only.

The sentiment of the people can be analyzed using Sentiment Analysis (SA) [9]. According to [10], [11] and [12], Sentiment Analysis is a process of extracting the user's emotions, feelings, or opinion and classify them into {positive, negative or neutral}. It is a field of natural language processing (NLP) [13]. A number of studies explained that due to the availability of search engines and social media websites like Google, Twitter and Facebook, people have access to a massive amount of data than ever before [10], [14], [12], [15], [16], [17]. One can find a lot of reviews and comments from all over the world regarding any particular topic, product or service as people use such social media platforms on a daily basis to express their opinion using their smartphone applications. Twitter offers a smartphone application for all major

platforms and has become a primary source of data for many studies. Hence, in this study, Twitter is used as a data source, and SA is applied on extracted text to detect people’s sentiment regarding the impact of technology on employment. After the sentiment is detected from the semi-structured and preprocessed text, it is used to build a machine learning classifier which can classify any new unseen text into {positive, negative or neutral}.

There are many studies using machine learning, rule-based and lexicon-based approaches of sentiment analysis in different domains such as [11], [15], [18], [19], [20] and [21] but machine learning approach shows great potential.

This study has two main objectives, 1) to preprocess text collected from Twitter concerning the impact of technology on employment and analyze its sentiment, 2) to evaluate the performance of machine learning Naïve Bayes classifier on the text.

The paper’s organization is as follows: In Section 2, related works are summarized. In Section 3, the proposed method is explained for preprocessing, analyzing sentiment, and building machine learning classifier. Section 4 shows results and analysis. Finally, the work is concluded.

2 Related Work

Social media sites like Twitter allows users to create or share content anytime and from anywhere in real life and time. It allows sensing some occurrence or trend or changes in lives. For example, when a product is launched, people post about it on social media. One can fetch that text and apply sentiment analysis to find out whether majority people are either happy and satisfied with the product or either they are neutral which means not happy nor sad or they are sad and unsatisfied. There is rarely any research on the impact of technology on employment, but researchers have used sentiment analysis in other domains. According to the studies [22] and [23], Sentiment Analysis has applications in almost every domain such as it allows hospitals to monitor social media websites in real-time so that they can act accordingly to improve health services. It also can be used in stock picking, which eventually will lead to superior returns. It can be used to classify the review of any product into positive or negative [24]. Figure 1 shows the general steps in sentiment analysis that are followed to analyze text for the sentiment.

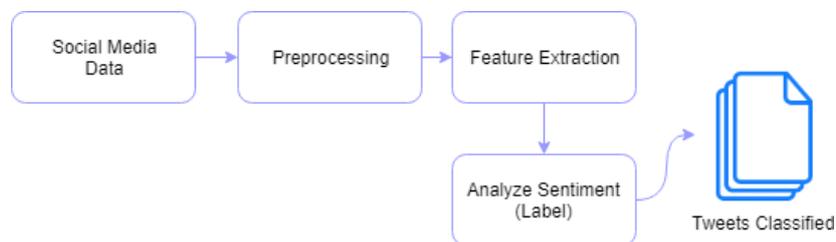


Fig. 1. Sentiment Analysis Flow

Sentiment Analysis can also be utilized in order to monitor the reputation of any specific brand on social media platforms. It can also help campaign managers to track how different voters feel about issues or how they relate to the actions and speeches of different candidates as also highlighted by [25]. Sentiment analysis can also be used to understand customer needs on any specific time and place. It can enlighten the customer's demand, as mentioned by [26].

2.1 Rapid miner

The first step to analyze the social media text is to consider a tool which can efficiently extract, manage, and process the large text. There are more than twenty tools available according to [15], which supports sentiment analysis, and Rapid Miner is one of the most popular tool [24], which is used in this study. Rapid Miner has features to extract text from Twitter, preprocess it, analyze its sentiment, build a machine learning classifier, and evaluate the performance. Many studies have used Rapid Miner to apply sentiment analysis, such as [15] and [24].

2.2 Sentiment analysis and classification

Studies such as [10] have worked on the comparison of different techniques and methods that are used in Sentiment Analysis. According to the authors, the Sentiment Analysis is another kind of text classification that classifies the text by its orientation of the opinions. [10] defined sentiment analysis as It is a process to detect the polarity of any given text and determines that if the given text is {neutral, positive, or negative}. Sentiment analysis has mainly three approaches, i.e., Machine learning, Rule-based, Lexicon-based [10].

Another study [11] used a machine learning approach and applied Support Vector Machine (SVM) with domain-specific lexicons on a corpus of 1940 reviews. The maximum achieved accuracy was 78.05% after testing the model on 41 reviews. An experiment was conducted using Rapid Miner by [13] in order to derive the sentiments from tweets and compared the accuracy of different algorithms. The study [15] used training data of 400+ tweets to apply classification model using SVM, Decision Tree, and Naïve Bayes (NB). SVM shows 79.08% accuracy while the Decision Tree shows 75.16% and NB shows 76.47% accuracy.

Another research [16] was conducted on a rule-based approach on 200 financial news articles and achieved 75.6% accuracy. An experiment on sentiment analysis using rule-based approach was conducted by [17] on 4,45,509 product reviews and reported 72.04% accuracy.

[20] applied a lexicon-based approach to 6,74,412 tweets and achieved 73.5% accuracy. Similarly, [21] also experimented with lexicon approach on 3,08,316 tweets, and the accuracy achieved is 82% in the multi-class classification with slangs.

Table 1. Summary of methods used & accuracies achieved

Study	Approach	Dataset (Rows)	Accuracy
[11]	Machine Learning	1940	.78.05%
[15]		400+	79.08%
[18]	Rule-Based	200	75.60%
[19]		4,45,509	72.04%
[20]	Lexicon-Based	6,74,412	73.50%
[21]		3,08,316	82%

Table 1 shows the fact that despite having smaller dataset size, machine learning approach achieves decent accuracy.

Table 2 shows the strengths and weaknesses of each of the SA approach discussed

Table 2. Strength & weaknesses of sentiment analysis approaches

Approaches	Classification	Advantages	Disadvantages
Machine Learning	Supervised/Unsupervised	Dictionary is not required High classification accuracy	Classifier once trained on the text of one specific domain cannot work for other domains
Rule-Based	Supervised/Unsupervised	91% performance accuracy at review level & 86% at the sentence level Sentence level classification works better than word level	Accuracy/Efficiency depends on defined rules
Lexicon-Based	Unsupervised	Learning procedure & labeled data is not required	Powerful linguistic resources are required

Table 2 depicts the fact that the machine learning approach offers a high accuracy as also shown by Table 1, and it does not require a dictionary every time to predict. Hence this study is focused on this approach. There is rarely any study on the impact of technology on employment from a sentiment analysis perspective, but researchers such as [1] have proposed a theoretical model to assess how artificial intelligence is shaping the job market and replacing humans. Hence, there is an urgent need to analyze people’s sentiments regarding the situation.

3 Proposed Method

In this study, the content related to the impact of technology on employment is identified and analyzed to get user sentiments from Twitter. The study also focused on training Naïve Bayes machine learning classifier to classify the content according to the user’s sentiment. Figure 2 shows the flow and significant steps of the proposed method.

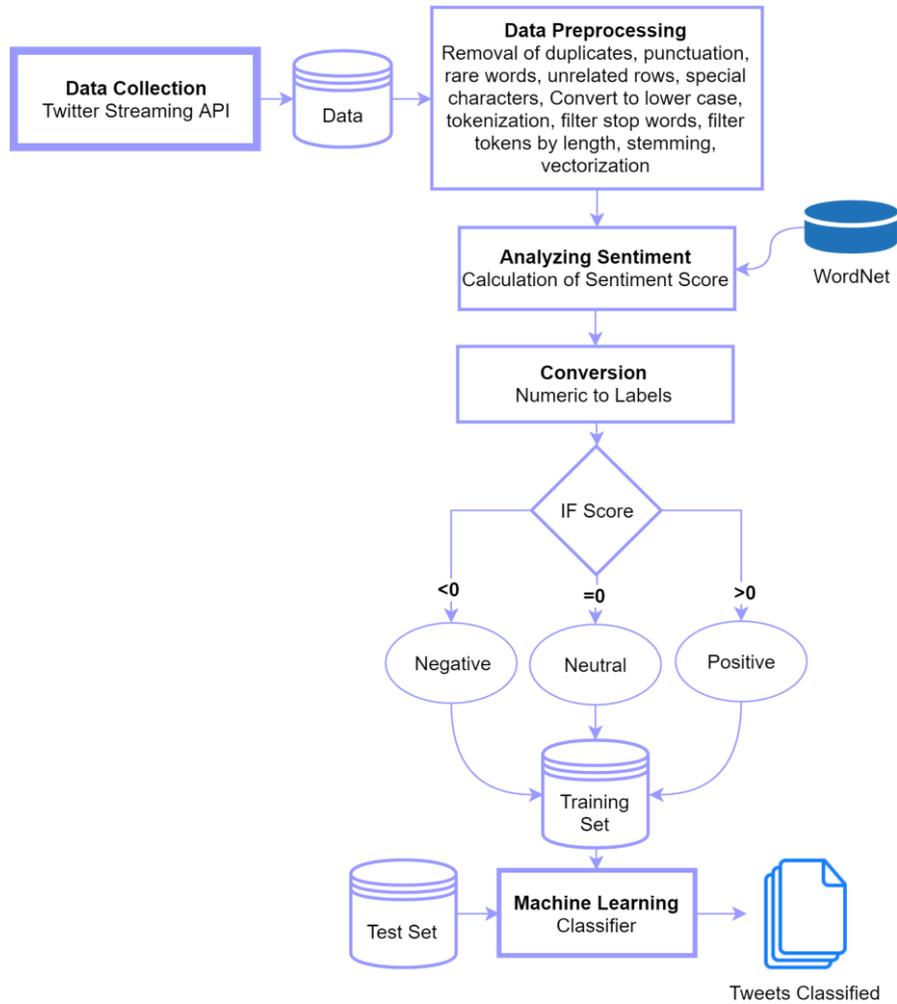


Fig. 2. The flow of the proposed method

3.1 Data collection

In order to collect data, first, there is a need to identify the keywords which can be used to search text on Twitter. The right selection of keywords fetches useful and closely related text, i.e., the impact of technology on employment and wrong keyword selection results in unrelated and useless text being fetched. These keywords are sometimes known as seed words. The process to choose the seed words for this study is to crawl the most related documents, i.e., the World Economic Forum Report about the job outlook in 2022 [5].

Following are the seed words that were selected by frequency and number of records fetched against them.

Table 3. Selected seed words

No	Seed Words	Records Fetched
1	AI automation replace	836
2	Robots taking over	340
3	AI taking over	251
4	Machines taking over	100
5	2022 jobs	479
6	AI replace human	449
7	Automation jobs	787
8	Unemployment technology	300
9	Technology taking over	546
10	Technology replace human	201
Total		4289

Table 3 shows seed words and the number of records fetched against those using Rapid Miner. There was a total of 4289 records initially before the text preprocessing stage. This number also includes duplicate records and some useless or unrelated comments. The text collection was realized using Rapid Miner by considering three filters, i.e., 1) recent or popular, 2) English only and 3) starting from any date till March 10, 2019.

3.2 Data preprocessing

Before analyzing the sentiment, the text was preprocessed.

First of all, all the duplicate rows were deleted leaving 1074 records out of 4289 initially. Then, punctuations were removed such as commas, full stops, periods, question marks, exclamation point, semicolon, colon, hash, hyphen, parenthesis, brackets, braces, apostrophe, quotation marks, and ellipsis. After that, all rare words were removed because they were very infrequent, and the association between infrequent and frequent words is mostly dominated by noise.

After this, unrelated rows were removed. When Twitter is searched using seed words, most of the time, there is a possibility of unrelated rows to appear in search results. Such rows can be slightly related to the seed word that is being used to search tweets, but those are not related to the extent that it gives any meaning in a specific context. Hence, all such rows were removed.



Fig. 3. Data preprocessing steps

After removing the punctuations, there is still a possibility of any special character that is left there which does not count under punctuation hence all such special characters were also removed including the user names of the users posting text on Twitter using regular expressions.

After completing the first five steps, as shown in Figure 3, the text was transformed into lower case. Then, the text was tokenized to create the sequence of tokens. Then, an English stop words dictionary is used to remove all stop words. After that, the tokens are filtered by their length. Any token that contains more than 20 characters was filtered. Finally, the stemming was used to fix affixes such as suffixed, prefixes, infixes, and circumfixes.

Finally, the preprocessed text is a large dataset which needs to be broken into smaller pieces so that it can be solved mathematically. The text needs to be converted to its integer representation, and for that purpose, the TF-IDF algorithm was used. It calculates term frequency and inverse document frequency (TF-IDF) for each token to create word vectors [27].

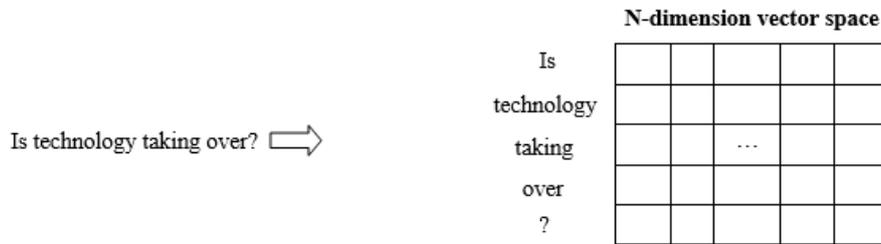


Fig. 4. N-dimension vector space

TF-IDF algorithm works on the following mathematical equation [28].

$$W_{i,j} = tf_{i,j} \times \log \left(\frac{N}{df_i} \right) \tag{1}$$

TF-IDF is a combination of two different terms, and both terms need to be calculated to find the value of TF-IDF, as shown in Equation 1. For analyzing the sentiment, first, the tweets need to be converted into word vectors.

After applying TF-IDF, the integer representation of each row can be achieved, as shown in Table 4 below.

Table 4. Word vector example

	also	love	programming
Love programming	0.000000	0.707107	0.707107
Programming also love	0.704909	0.501549	0.501549

Table 4 shows the word vectors for two rows. It is created for every word and row in the dataset so that the text can be processed mathematically for sentiment analysis and machine learning classifier.

3.3 Analyze sentiment

After successfully preprocessing the text, it was ready to be analyzed for the sentiments. A lexical English database called WordNet was used to analyze the sentiment as it has been already used by many studies [29], [30].

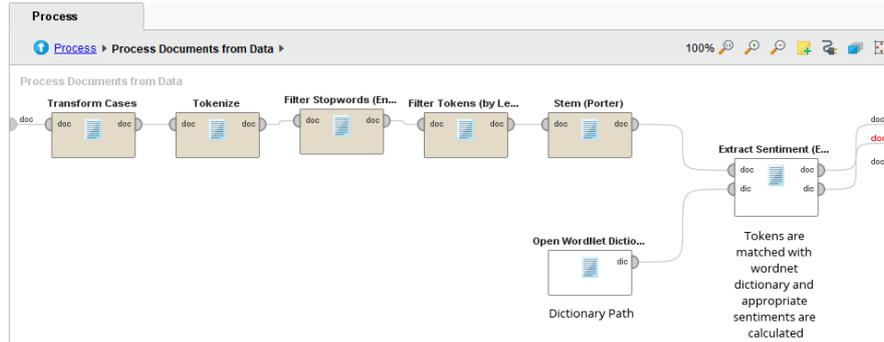


Fig. 5. Analyzing sentiment using WordNet in Rapid Miner

Figure 5 shows the steps performed after the preprocessing stage. WordNet can provide the sentiment of the text in either negative numeric values or positive numeric values including zero where negative value means “Negative” sentiment, zero value means “Neutral,” and a positive value means “Positive” sentiment as shown in Table 5 below.

Table 5. Selected seed words

Seed Words	Score	Label
i tell people on here at least monthly that majority of the jobs will be gone in the near future	-0.49967	Negative
damn evil robots taking over	-0.25852	Negative
what oh god artificial intelligence is taking over the world	-0.375	Negative
not too worried yet about robots taking over musicians’ jobs	0	Neutral
robots may not be taking over jobs but they are changing them	0	Neutral
not worried about elon musks vision of terminator like robots taking over humanity	0	Neutral
robots taking over to help medical research	0.2	Positive
those warning about a machine takeover typically assume artificial intelligence will develop super intelligence	0.044886	Positive
i shall worry about artificial intelligence taking over when siri can actually read what I ask her about	0.096715	Positive

Table 5 shows the converted values of the analyzed sentiment from numeric into three class labels, i.e., {positive, negative, neutral} so that a supervised machine learning classifier such as Naïve Bayes can be trained.

3.4 Building machine learning classifier

After getting the labeled text, now a supervised machine learning classifier, i.e., Naïve Bayes can be trained so that it can classify any unseen text into {positive, negative, neutral}. The preprocessed text contains a total of 1074 rows. 100 rows were kept as a test set for later use leaving 974 rows for the training and validation part. 974 rows were divided into training and validation according to 80:20 ratios, as suggested by the famous Pareto Principle [31]. Once the classifier is trained and validated, the test set is used to test the model.

Naïve Bayes classifier is based on the Bayesian Theorem that follows a naïve assumption about the features to be independent of each other. It assigns a document (d) to the class (c) which maximizes the $P(c|d)$ with the help of Bayes rule [11].

$$P(c|d) = \frac{P(c)P(d|c)}{P(d)} \tag{2}$$

Naïve Bay is a high-bias, but low variance classifier and a good model can be built using it with even a small data set.

In Rapid Miner, some operators can be used to build a machine learning classifier, as shown in Figure 6. The sequence of steps starts from the left and goes towards the right. Figure 6 illustrates the process to build a machine learning classifier in Rapid Miner. First, the classifier is trained, then it is validated, and finally, it is tested using Test Set.

First of all, an excel spreadsheet is provided as a training set which has two columns, i.e., 1) text and 2) polarity. The “text” column contains the tweets and polarity contains the labels such as {positive, negative, neutral}.

Secondly, the data is preprocessed and converted into word vector using TF-IDF, as shown in Figure 4. Then, “Set Role” operator is used to set the target variable, i.e., “Polarity” and “Select Attributes” is used to filter out the rows with missing labels.

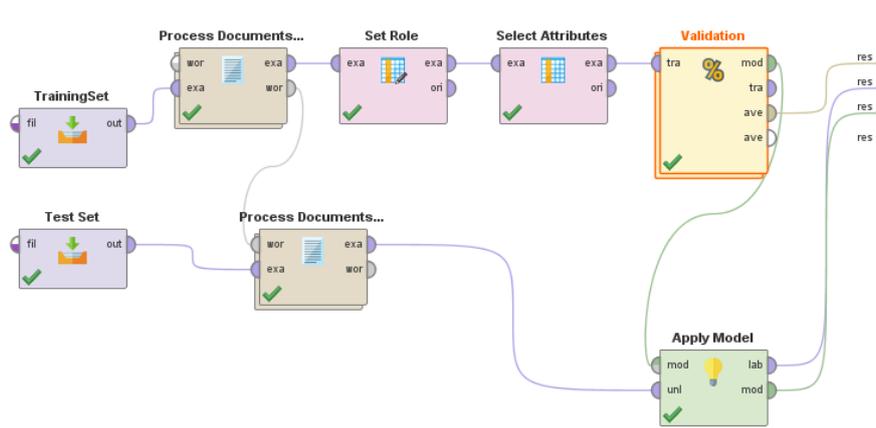


Fig. 6. Machine learning classification process

The “Validation” operator is used to apply the Naïve Bayes (NB) classifier and test its performance. At this stage, the training set is split into two groups using 80:20 ratios, i.e., “training” and “validation.” The training set is used by the classifier to learn patterns and relationships in order to fit parameters, i.e., weights but this approach usually overfit the data, and that is why validation set is used to adjust the hypermeters to avoid overfitting.

Once the classifier is trained, and validation has taken place, another excel spreadsheet is provided as a test set which only has one column, i.e. “text.” This file does not have the labels or polarity because that is what machine learning classifier is going to predict. After training and validation, the test set, which is independent of the training set, is used to assess the performance of the classifier. The test set is also passed through the preprocessing stage to make it word vectors using TF-IDF as it was done for training set then it is passed to “Apply Model” operator to connect it with the trained and validated Naïve Bayes (NB) classifier so that it can be assessed using the test set.

The evaluation of the classifier can be done using evaluation metrics such as accuracy, precision, and recall.

The accuracy of classifier reports that how often the classifier classifies something correctly. To calculate accuracy, true positive (TP), true negative (TN), false positive (FP) and false-negative (FN) values are used as shown below in Equation 3

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

The precision reports that when the classifier predicts something positive, how often it predicts it correctly. To calculate recall, true positive (TP) and false positive (FP) values are used as shown below in Equation 4

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

The recall reports that when the result is positive, how often the classifier predicts it correctly. To calculate recall, true positive (TP) and false-negative (FN) values are used as shown below in Equation 5

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

Naïve Bay is a high-bias, but low variance classifier and a good model can be built using it with even a small data set.

4 Results and Analysis

After collecting and preprocessing the tweets from Twitter regarding the impact of technology on employment, its sentiment was analyzed using WordNet, which is a lexical English database that was used to analyze the sentiment. There was a total of 974 rows in the dataset for which the sentiments were analyzed as shown in Figure 7 below

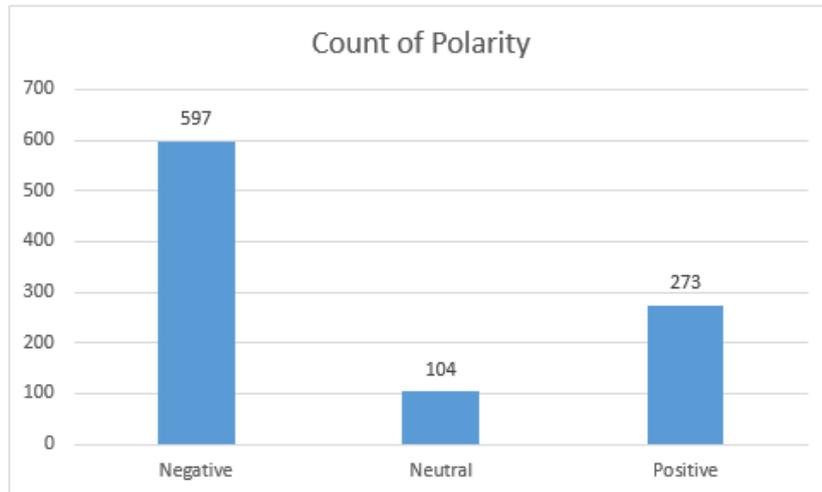


Fig. 7. Count of polarity after analyzing the sentiment

Figure 7 shows the labeled values by WordNet for each category {positive, negative, neutral} and depicts the fact that 61.3% of the users on Twitter whose data was collected had negative sentiment regarding the impact of technology on employment. While on the other hand, 28.03% of the users had a positive sentiment, and only 10.67% of users were neutral. That means the majority of the people are of the view that technology is going to impact their employment negatively and they are worried about the recent advancements in technology especially, artificial intelligence, automation, machines, and robots.

Now the labeled text is available; hence, a machine learning classifier can be built so that it can classify new unseen text into {positive, negative, neutral} categories.

A dataset containing 974 rows is used with a split of 80:20 ratios, where 80% is for training the Naive Bayes (NB) and Support Vector Machine (SVM) classifiers and remaining 20% is for validation. The test set has 100 rows using which it is tested.

Table 6 shows 87.18% overall accuracy with acceptable recall and precision values.

Table 6. Naive Bayes classifier confusion matrix

Accuracy 87.18%	True Negative	True Neutral	True Positive	Class Precision
Pred. Negative	83	0	0	100.00%
Pred. Neutral	16	59	2	76.62%
Pred. Positive	7	0	28	80.00%
Class Recall	78.30%	100.00%	93.33%	

The Naïve Bayes has classified the test set (100 rows) into {negative, neutral, positive} sentiments as shown below in Figure 8

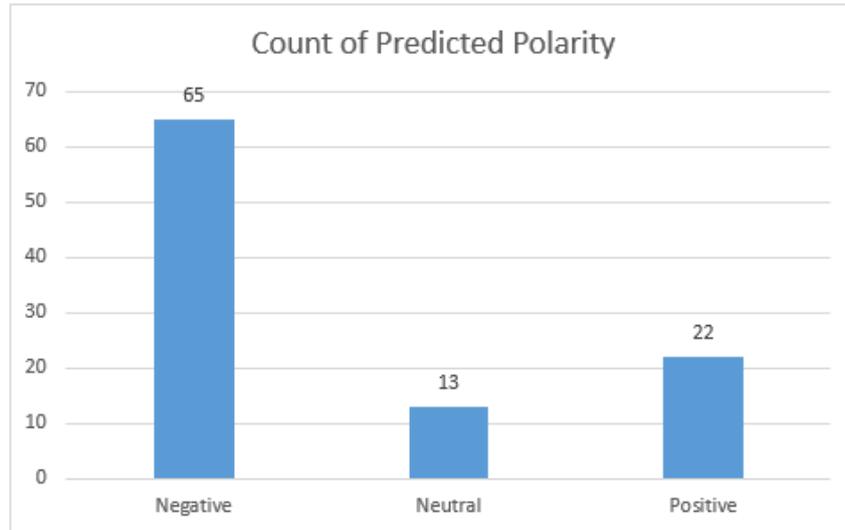


Fig. 8. Classification of 100 rows (Test Set) using Naïve Bayes

Figure 8 depicts the fact that the results are similar to what was observed in the dataset before building classifier in Figure 7. Most of the users have negative sentiment regarding the impact of technology on employment as 65% of tweets are classified as “Negative” which is very close to the percentage value of 61.3% as shown in Figure 7.

The same experiment was repeated, but this time with Support Vector Machine as it is another popular and very frequently used classifier [32]. By using SVM, the following measures are achieved as shown in Table 7

Table 7. Naive Bayes & SVM

Model	Naïve Bayes	SVM
Accuracy	87.18%	82.05%

Table 7 shows that Naïve Bayes has achieved better accuracy than SVM.

The study has found that 65% of the people for which this experiment is conducted holds negative sentiment regarding the impact of technology on employment hence, once it is identified that a large number of people are of the view that technology is going to take over their jobs, following measures can be taken:

- Such a group of people can improve their skillset by self-learning to match the changing requirements of the industry. Learning through the MOOCS and attending the conferences on emerging technologies can help them to remain updated.
- Human Resource department of any organization can conduct a similar study to find out the group which is severely affected among employees and must arrange training and workshop series to improve their skills to prepare them for modern tasks and needs of the organization.

- Organizations embracing automation and technologies like AI should conduct similar studies and re-consider their policies that up to what extent automation is beneficial to the aim and objectives of the organization and society. There must be a balance while adapting technology based on their advantages and disadvantages. Each organization holds a social responsibility, and it must take care of its employees.

It is time for the individuals and organizations to realize while reaping the benefits of technology that it is a double-edged sword as it has the ability to liberate and enslave.

5 Conclusion and Future Work

A system is proposed to analyze the sentiment of the people regarding the impact of technology on their employment and to build a machine learning classifier using Naïve Bayes in order to classify any unseen text of this context easily. The Rapid Miner is used to collect, manage, preprocess, and analyze the sentiments along with the WordNet dictionary. Furthermore, it was also used to build machine learning classifier. The text was collected from Twitter using seed words which are either recent or popular from any date till March 10, 2019, and is in the English language.

The study has found that the majority of the people whose tweets were collected and analyzed have negative sentiments regarding the impact of technology on employment and advancements in technologies like Artificial Intelligence, Automation, and Robotics.

There is rarely any research that has contributed to this domain. Due to limited availability of data in this domain, there is a limitation of labeled training data for building machine learning classifier as such classifiers usually perform better when fed with a large amount of data. Secondly, the WordNet dictionary also has limitations as it struggles to distinguish simple and multi-word units. Despite these limitations, the model has achieved 87.18% accuracy using Naïve Bayes classifier.

People having negative sentiments must be encouraged to learn new skills that can keep them relevant in the 21st century of automation so that they can be saved from the effects of structural unemployment. Furthermore, more data can be collected in the future to increase the training set so that machine learning classifier can offer even better accuracy. Finally, instead of using WordNet dictionary for analyzing sentiments, other approaches can be used such as SentiWordNet and other automated methods being offered by the developers such as Aylie.

6 References

- [1] M. H. Huang and R. T. Rust, "Artificial Intelligence in Service," *J. Serv. Res.*, vol. 21, no. 2, pp. 155–172, 2018.
- [2] G. Su, "Unemployment in the AI age," *AI Matters*, vol. 3, no. 4, pp. 35–43, 2018. <https://doi.org/10.1145/3175502.3175511>
- [3] N. J. Nilsson, "Artificial intelligence, employment, and income," *Hum. Syst. Manag.*, vol. 5, no. 2, pp. 123–135, 1985.

- [4] D. Rotman, “How Technology Is Destroying Jobs,” *Technol. Rev.*, vol. 16, no. August, pp. 28–35, 2013.
- [5] I. Report, *The Future of Jobs Report 2018 Insight Report* Centre for the New Economy and Society. 2018.
- [6] Steve Crabtree, “In U.S., Depression Rates Higher for Long-Term Unemployed | Gallup,” *Well-Being*, 2014. [Online]. Available: <https://news.gallup.com/poll/171044/depression-rates-higher-among-long-term-unemployed.aspx>. [Accessed: 08-Dec-2018].
- [7] S. Boseley, “Unemployment causes 45,000 suicides a year worldwide, finds study,” *The Guardian*, 2015.
- [8] C. Nordt, I. Warnke, E. Seifritz, and W. Kawohl, “Modelling suicide and unemployment: A longitudinal analysis covering 63 countries, 2000-11,” *The Lancet Psychiatry*, vol. 2, no. 3, pp. 239–245, 2015. [https://doi.org/10.1016/s2215-0366\(14\)00118-7](https://doi.org/10.1016/s2215-0366(14)00118-7)
- [9] F. Colace, M. de Santo, and L. Greco, “Safe: A sentiment analysis framework for e-learning,” *Int. J. Emerg. Technol. Learn.*, vol. 9, no. 6, pp. 37–41, 2014. <https://doi.org/10.3991/ijet.v9i6.4110>
- [10] M. D. Devika, C. Sunitha, and A. Ganesh, “Sentiment Analysis: A Comparative Study on Different Approaches,” *Procedia Comput. Sci.*, vol. 87, pp. 44–49, 2016. <https://doi.org/10.1016/j.procs.2016.05.124>
- [11] C. Bhadane, H. Dalal, and H. Doshi, “Sentiment analysis: Measuring opinions,” *Procedia Comput. Sci.*, vol. 45, no. C, pp. 808–814, 2015. <https://doi.org/10.1016/j.procs.2015.03.159>
- [12] D. M. E. D. M. Hussein, “A survey on sentiment analysis challenges,” *J. King Saud Univ. - Eng. Sci.*, vol. 30, no. 4, pp. 330–338, 2018.
- [13] M. AbdelFattah, D. Galal, N. Hassan, D. Elzanfaly, and G. Tallent, “A Sentiment Analysis Tool for Determining the Promotional Success of Fashion Images on Instagram,” *Int. J. Interact. Mob. Technol.*, vol. 11, no. 2, p. 66, 2017. <https://doi.org/10.3991/ijim.v11i2.6563>
- [14] S. Sun, C. Luo, and J. Chen, *A review of natural language processing techniques for opinion mining systems*, vol. 36. Elsevier B.V., 2017.
- [15] V. Vyas and V. Uma, “An Extensive study of Sentiment Analysis tools and Binary Classification of tweets using Rapid Miner,” *Procedia Comput. Sci.*, vol. 125, pp. 329–335, 2018. <https://doi.org/10.1016/j.procs.2017.12.044>
- [16] K. Ravi and V. Ravi, *A survey on opinion mining and sentiment analysis: Tasks, approaches and applications*, vol. 89, no. June. Elsevier B.V., 2015. <https://doi.org/10.1016/j.knosys.2015.06.015>
- [17] N. Öztürk and S. Ayvaz, “Sentiment analysis on Twitter: A text mining approach to the Syrian refugee crisis,” *Telemat. Informatics*, vol. 35, no. 1, pp. 136–147, 2018. <https://doi.org/10.1016/j.tele.2017.10.006>
- [18] L. I. Tan, W. S. Phang, K. O. Chin, and P. Anthony, “Rule-Based Sentiment Analysis for Financial News,” *Proc. - 2015 IEEE Int. Conf. Syst. Man, Cybern. SMC 2015*, pp. 1601–1606, 2016. <https://doi.org/10.1109/smc.2015.283>
- [19] C.-S. Yang and H.-P. Shih, “A Rule-Based Approach for Effective Sentiment Analysis,” *PACIS 2012 Proc.*, 2012.
- [20] C. Kaushik and A. Mishra, “A Scalable, Lexicon Based Technique for Sentiment Analysis,” *Int. J. Found. Comput. Sci. Technol.*, vol. 4, no. 5, pp. 35–56, 2014.
- [21] M. Z. Asghar, F. M. Kundi, A. Khan, and S. Ahmad, “Lexicon - Based Sentiment Analysis in the Social Web,” *J. Basic. Appl. Sci. Res*, vol. 4, no. 6, pp. 238–248, 2014.
- [22] M. T. Khan and S. Khalid, “Sentiment Analysis for Health Care,” *Int. J. Priv. Heal. Inf. Manag.*, vol. 3, no. 2, pp. 78–91, 2015.
- [23] R. Feldman, “Techniques and applications for sentiment analysis,” *Commun. ACM*, vol. 56, no. 4, p. 82, 2013.

- [24] P. S. Chauhan, “Opinion Mining and Sentiment Analysis,” 2017.
- [25] A. P. Singh, A. Malik, and D. Kapoor, “Sentiment Analysis on Political Tweets,” pp. 359–361, 2016.
- [26] O’Connell Brian, “How Sentiment Analysis and Data Analysis Can Improve Your Sales,” Business News Daily, 2017. [Online]. Available: <https://www.businessnews-daily.com/10018-sentiment-analysis-improve-business.html>. [Accessed: 09-Dec-2018].
- [27] S. Qaiser and R. Ali, “Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents,” *Int. J. Comput. Appl.*, vol. 181, no. 1, pp. 25–29, 2018. <https://doi.org/10.5120/ijca2018917395>
- [28] Diggity Matt, “TF*IDF for SEO: The Ultimate Beginner to Advanced Guide,” 2019. [Online]. Available: <https://diggitymarketing.com/tfidf-for-seo/>. [Accessed: 29-Mar-2019].
- [29] A. Srivastava, V. Singh, and G. S. Drall, “Sentiment Analysis of Twitter Data,” *Int. J. Healthc. Inf. Syst. Informatics*, vol. 14, no. 2, pp. 1–16, 2019.
- [30] S. Poria, A. Gelbukh, E. Cambria, P. Yang, A. Hussain, and T. Durrani, “Merging SenticNet and WordNet-Affect emotion lists for sentiment analysis,” *Int. Conf. Signal Process. Proceedings, ICSP*, vol. 2, pp. 1251–1255, 2012. <https://doi.org/10.1109/icosp.2012.6491803>
- [31] Gulipalli Pradeep, “The Pareto Principle for Data Scientists.” [Online]. Available: <https://www.kdnuggets.com/2019/03/pareto-principle-data-scientists.html>. [Accessed: 29-Mar-2019].
- [32] A. Abubakar et al., “A Support Vector Machine Classification of Computational Capabilities of 3D Map on Mobile Device for Navigation Aid,” *Int. J. Interact. Mob. Technol.*, vol. 10, no. 3, p. 4, 2016. <https://doi.org/10.3991/ijim.v10i3.5056>

7 Authors

Shahzad Qaiser is a Lecturer at department of Computer Science, Capital University of Science and Technology (CUST), Islamabad, Pakistan. shz.qais@gmail.com.

Nooraini Yusoff is an Associate Professor at Institute for Artificial Intelligence and Big Data (AIBIG), Universiti Malaysia Kelantan, City Campus, 16100 Kota Bharu, Kelantan, Malaysia. nooraini.y@umk.edu.my.

Farzana Kabir Ahmad is a Senior Lecturer at School of Computing, UUM College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia. farzana58@uum.edu.my.

Ramsha Ali is from School of Quantitative Sciences, UUM College of Arts and Sciences, Universiti Utara Malaysia, 06010 UUM Sintok, Kedah, Malaysia. ramshaali47@gmail.com.

Article submitted 2019-04-04. Resubmitted 2019-08-05. Final acceptance 2019-09-01. Final version published as submitted by the authors.