Contents

[Pre-Processing of Text 2](#_heading=h.gjdgxs)

[Maltese Targeted Systems 3](#_heading=h.30j0zll)

[Lexicon Based Approaches 4](#_heading=h.1fob9te)

[Machine Learning approaches 7](#_heading=h.3znysh7)

[Hybrid solutions 11](#_heading=h.2et92p0)

[List of approaches 12](#_heading=h.tyjcwt)

The aim of this chapter is to provide an overview of the various approaches that other researchers have taken in order to solve the problem of sentiment analysis. Since such approaches targeted towards the Maltese language are few, systems for other languages were studied in order to help design a better novel system targeted towards the Maltese language. The research carried out mainly focuses on analysing the pre-processing techniques used and proposed architectures employed in previous approaches. The systems covered in this section include both lexicon-based systems as well as machine-learning based systems in order to cover all possible solutions towards this problem.

# Pre-Processing of Text

The first step towards solving the problem of sentiment analysis is to compile a text corpus, composed of a structured set of texts. Pre-processing of text is done to bring the required text into a form that is predictable and analyzable for a particular task. In most scenarios, the required text is collected from online sources where reviews or opinionated text can be found. Once a substantial amount of data is collected, it undergoes pre-processing. Pre-processing of the text entails ‘cleaning’ any redundant and uninformative data from that collected and preparing the text for the classification stage. This step is common to both lexicon based, and machine-learning based approaches, excluding deep learning systems, which handle preprocessing without any human intervention. Numerous amounts of preprocessing techniques were mentioned in the systems researched.

The data used in [Neha, H. Gupta, S. Pande, A. Khamparia, V. Bhagat, and N. Karale, “Twitter Sentiment Analysis Using Deep Learning,” in *Materials Science and Engineering*, 2021, vol. 1022.] is gathered using the twitter API, as this data was prepared for the neural network. In this case, it was sufficient to delete unnecessary material such as emojis, unique characters and any extra whitespaces. In addition to this, several repeated tweets and tweets containing less than three characters were removed. Similar data pre-processing was applied in [S. Xiao*et al*, "Sentiment Analysis for Product Reviews Based on Deep Learning," *Journal of Physics: Conference Series,*vol. 1651, *(1),*2020. Available: https://ejournals.um.edu.mt/login?url=https://www-proquest-com.ejournals.um.edu.mt/scholarly-journals/sentiment-analysis-product-reviews-based-on-deep/docview/2513091833/se-2?accountid=27934. DOI: <http://dx.doi.org.ejournals.um.edu.mt/10.1088/1742-6596/1651/1/012103>.]

Pre-processing of foreign text also supports the aforementioned techniques. For instance, the approach used in [C. Frans, Siti Mariyah, and Setia Pramana, “Sentiment analysis: a comparison of deep learning neural network algorithm with SVM and naϊve Bayes for Indonesian text,” *Journal of Physics: Conference Series*, vol. 971, p. 012049, 2018, doi: 10.1088/1742-6596/971/1/012049.] involves the following five main steps: Case Folding, Word Tokenizing, Stop Word Removal, Stemming, Text Formalization. Of note, the importance of checking for any typos in the given text is also highlighted. Indeed, one such small mistake can alter the sentiment of a word and so this can have a negative impact on the performance.

However, this does not come without its flaws. As [ElBeltagy, Samhaa R, T. Khalil, A. Halaby, and M. Hammad, “Combining Lexical Features and a Supervised Learning Approach for Arabic Sentiment Analysis,” in *Computational Linguistics and Intelligent Text Processing*, 2018, pp. 307–319.] and [A. C. Forte and Brazdil, Pavel B, “Determining the Level of Clients’ Dissatisfaction from Their Commentaries,” *Computational Processing of the Portuguese Language*, vol. 9727, no. 016–0621, pp. 74–85, 2016, DOI: 10.1007/9783319415529\_7.] point out, a problem arises following data stemming. A. C Forte et al. discovered that since the work done is based on foreign languages, when stemming was applied to the text, the meaning of a word was also changed and therefore, stemming was discarded. This problem was overcome through the use of lemmatization, which selects a representative form (e.g. infinitive) from a given set of words having the same root. To carry out lemmatization, a dictionary-based tool was utilised, which involves a large set of dictionaries built using large datasets. An additional advantage of this is that the tool allows for the addition of manual entries. Base forms of lexicon entries were added to a “stem list” to ensure that any word matching those on the list, does not get stemmed beyond its base form. Both the tweets/texts and lexicon entries are lemmatized and stemmed using this tool, prior to any matching steps. When a match occurs between text in the input and a lexicon entry, a unique term is added to the input text immediately after the matching sentiment term. This depends on whether the match was with a positive or negative entry.

To further prepare the required data for the problem of sentiment analysis, several works [V. S. Lakshmi, K Janan, J. P. S. Joshua, and M. Sharoz, “Predicting supervised machine learning performances for sentiment analysis using contextual based approaches,” *Journal of Physics: Conference Series*, vol. 1916, Art. no. 1, 2021, doi: 10.1088/1742-6596/1916/1/012117.][Nichole thesis] decided to use POS tags. In linguistics, part-of-speech tagging is used, which involves assigning a word in text to a particular “category” of speech, based on both its definition and its context. A simplified form of this is commonly taught to school-age children, when categorising words as nouns, verbs, adjectives, adverbs, etc. However, in the work done by Nichole et al., the POS tagger proved to be uninformative in the case of their corpus. Hence, what is certain about text pre-processing is that there are several ways to go about it, however a system that is practically faultless is yet to be designed.

# Features

The work in [[1]R. Ahuja, A. Chug, S. Kohli, S. Gupta, and P. Ahuja, “The Impact of Features Extraction on the Sentiment Analysis,” *Procedia Computer Science*, vol. 152, pp. 341–348, 2019, doi: 10.1016/j.procs.2019.05.008.] involves the use of 6 different classification algorithms applied on the SS-Tweet dataset considering two features (TF-IDF and N-Grams). Following sentiment analysis of several tweets, it was concluded that TF-IDF features produced better results (3-4%) when compared to N-Gram features. Therefore, this leads one to conclude that TF-IDF is the better choice, when compared to N-Gram, in terms of features relating to the use of machine learning algorithms for text classification.

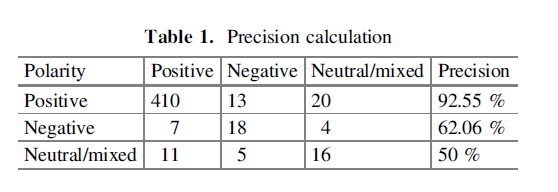
# Maltese Targeted Systems

Two projects currently exist that have attempted to design a sentiment analysis system targeted towards the Maltese language. Nichole et al. [Nichole thesis] managed to produce a machine learning, context-based system capable of classifying a given piece of text written in Maltese as positive, negative, or neutral. The researchers incorporated context into their approach by performing classification at a context-window level, allowing their system to not only identify the different sentiments expressed within the same sentence, but to also determine the polarity of that sentiment within the context of its surrounding words. The developed artefact is composed of two main components, both capable of performing preprocessing, feature identification and extraction, and classification of Maltese written text at context window level. The Custom Feature Component (CFC) follows the more traditional approach where preprocessing and feature identification is carried out manually, and the extracted hand-crafted features, including unigram value, POS tag, stem word value and negation presence, are passed on to more conventionally used algorithms including Naive Bayes, Maximum Entropy, SVM, a Decision Tree and a Random Forest classifier. On the other hand, the second component, known as the Unsupervised Feature Component (UFC), performs preprocessing and feature extraction without any human intervention, and passes on said features to classifiers based on deep learning architectures. Within this component a deep learning classier known as a Deep Belief Network was implemented. Through experimentation it was decided that a Random Forest classifier, trained using 80% of the original dataset with context windows composed of 4 words, yielded the optimal results. This setup achieved a maximum accuracy of 62.3%. By means of the tests carried out it was also concluded that the unigram and stem word values were the most informative features used within the CFC. The researchers also stated that the measures taken in an attempt to overcome the bias towards the negative class in the dataset were fruitless.

The work done in [dawson thesis] is a continuation of the work done by Sant et al. the researchers decided to make use of two algorithms that was used in the study done by Sant et al. and these are RF and SVM. In [Nichole thesis] the results were better as the dataset used was much larger. The accuracies produce in the work done by Camilleri et al. were very low as the dataset used for the training of the algorithms was very small and this heavily impacted the performance of such algorithms.

# Lexicon Based Approaches

Within this section a few lexicon-based techniques designed by other researchers to carry out the task of classification shall be reviewed. These include techniques which utilize scoring methods and lexicon-based knowledge graphs for determining the sentiment polarity of texts. The system designed by S. Zeb et al [S. Zeb, U. Qamar, and F. Hussain, “Sentiment analysis on user reviews through lexicon and rule-based approach,” in Web Technologies and Applications, 2016, vol. 9865, pp. 55–63.] makes use of a rule-based approach along with a lexicon-based approach in order to obtain a high accuracy sentiment classifier. Following pre-processing of the text, the polarity of the data is determined through a lexicon and by calculation of semantic orientation of the words. Patterns of two-word phrases are investigated, and they are then used as rules. These said rules check the polarity of the data depending on the sequence, whilst checking for adjectives and adverbs simultaneously. These patterns of tags for two-word phrases proved to be very useful in determining the polarity of the data, as the sequence or tags act as useful indicators of the sentiment in the review. The two-word phrases are extracted, and then the polarity in these phrases and the polarity through adjectives and adverbs were calculated. Each word in the review was matched with the frequent positive/negative words in the database. If the number of positive terms was greater than the negative, the review was considered positive. If the number of negative terms was greater than the positive, the review was considered negative. A third scenario is possible, whereby the number of positive and negative words is equal. In this case the review is considered to be neutral. Each possible word is checked through iterations, following which results are generated. This system managed to produce the following results:



Such a system presented researchers S. Zeb et al with its own set of challenges. One such challenge in particular, involved several reviews containing sarcastic and/or ironic terms. In addition to this, the presence of terms having dual meaning proved to be very difficult to analyse and so, this affected the efficiency of the developed algorithms. However, the system produced is a good starting point for a sentiment classifier. This actively highlights the need for more advanced algorithms, with the capability to specifically handle such terms, in order to produce better results.

Another lexicon-based approach was taken in [A. Sarigiannidis, P.-A. Karypidis, P. Sarigiannidis, and I.-C. Pragidis, “A novel lexicon-based approach in determining sentiment in financial data using learning automata,” in *Journal of Physics: Conference Series*, 2018, pp. 37–48.], however it focuses on supervised learning, rather than using a readily available lexicon as done in [S. Zeb, U. Qamar, and F. Hussain, “Sentiment analysis on user reviews through lexicon and rule-based approach,” in Web Technologies and Applications, 2016, vol. 9865, pp. 55–63.]. This system is able to create a new lexicon based on annotated textual data, and then applies that lexicon to determine the sentiment in new, not-annotated data. To build the automaton lexicon, the word polarity vector is initially defined. The higher the word polarity, the more positive the word is and vice versa. Every time a word is recorded in an annotated phrase, a probability reinforcement scheme is applied. The automaton lexicon is built using 4 main steps: pre-processing of data, creation of word polarity vector, training, and final output. Consider p(x) to define the word polarity vector of the initial training set. Assuming that m denotes the total number of words that appeared in the training set, the word polarity vector is initialized as follows:

p(i) = 1/m for each word i, 1 ≤ i ≤ m

to train the automaton lexicon. The word polarity function is updated accordingly, for each examined sentence, as follows:

p(x) = p(x) + R/L ・ p(x) for each word x found in the sentence (4)

p(y) = p(y) + S/(m − w) ・ p(y) for each other word y (5)

where R is the annotation score of the examined sentence, the parameter L is the convergence speed of the automaton, the parameter S is the produced overall increase of the sentence upon the polarity vector and the parameter w stands for the number of the words found in the specific sentence. The output of the automaton lexicon building process, is a custom lexicon which is composed of a set of words along with their polarity score.

After the system provides an output, normalisation is required since a polarity vector does not express negative values. Normalisation is also needed in order to apply the automaton lexicon to sentiment analysis determination in the range [−1, 1]. The normalized polarity vector (pN(x)) is defined as follows:

Where the smallest value in that range denotes the most negative recorded polarity value, and the largest value denotes the most positive recorded polarity value in the formed lexicon.

To make the extracted automaton lexicon more accurate, word reliability is introduced. Word reliability refers to the consistency a word has as the lexicon is formed. The frequency at which a word changes polarity is monitored. Words which change polarity frequently are regarded as ‘unreliable’ and vice versa. The word reliability score is determined as follows. To start off with, during the automaton lexicon building stage, the frequency of each word is calculated. Each time a word is recorded to change polarity, its unreliability score rises. This increase is higher (+1) when the word changes from positive to negative and vice versa, while it is lower (+0.5) when the word changes from neutral to positive/negative and vice versa. Then, depending on the obtained results, the proposed automaton lexicon attempts to determine the sentiment of four financial indices (oil price, euro Stoxx, European central bank and euro/dollar). A. Sarigiannidis et al concluded this to be quite accurate.

Since the proposed system is designed for the Maltese language, and systems that provide good results for this language are limited, sentiment analyses for foreign languages were analysed. The following work [A. C. Forte and Brazdil, Pavel B, “Determining the Level of Clients’ Dissatisfaction from Their Commentaries,” *Computational Processing of the Portuguese Language*, vol. 9727, no. 016–0621, pp. 74–85, 2016, DOI: 10.1007/9783319415529\_7.] analyses comments transcribed by assistants of a help-desk service of a Portuguese telecommunications company. Again, a lexicon-based approach was adopted, and it is based on the so-called “opinion words”. Here, the sentiment of a given comment is deduced by processing all words, and the sentiment value is recorded in cases of particular words appearing in the sentiment lexicon. The sentiment value of the whole text is consecutively determined by summing up all the sentiment values encountered. Researchers A. C. Forte et al have also catered for inaccuracies that occur when the system is applied to the Portuguese language, to improve its accuracy.

Not all terms in the given lexicon were useful and some of them were also affecting the performance negatively. Therefore, to resolve this, Backward Elimination method, which is similar to the feature selection method, was applied. This involves the processing of all terms in turn. At every step, the system determines the effect of eliminating a particular term on the overall system performance. If the performance decreased, the term is maintained, but otherwise, it is dropped. Other enhancements carried out to maximise performance include enhanced pre-processing, restricted stemming, rewriting of rules, removal of short texts and the construction of domain-specific stop words. In addition to the aforementioned, the designed system is also aimed to deal with polar expressions. Without any special treatment for such cases, the given result would be contradictory. Also, in order to account for cases whereby the word after a negation contained a positive sentiment, the inversion rule was introduced. This reversed the sentiment value of the associated expression. On the contrary, for cases in which the word after a negation had a negative sentiment, the inversion rule was corrected. Indeed, the results proved that such a change improved overall performance. Therefore, a list of amplifiers and attenuators was devised. When an amplifier is encountered and the subsequent is positive, 1 is added, else 1 is subtracted. When an attenuator is encountered and the subsequent is positive, 1 is added, else 1 is subtracted. The final version incorporating the above-mentioned enhancements was that of AffinPDO. It has an accuracy of 81.9% making it the best performing system out of the systems tested by the researchers.

On the conquest of achieving maximal performance and accuracy, A. C. Forte et al tested the application of off-the-shelf solutions in conjunction with lexicon Sentilex, available for Portuguese. This produced poor results, and therefore another set of techniques were applied, which involved translating the terms of the chosen lexicon (Affin) into Portuguese and then enriching it with domain-specific terms. This was found to be rather time-consuming, and it was concluded that the system had areas of improvement, nevertheless. Therefore, the possibility of generating an enriched lexicon through automated methods, rather than through a manual approach, was investigated. Indeed, it led to a marked improvement in performance. This is one of the reasons as to why the final system reached such an excellent performance both in terms of accuracy and mean cost.

Unfortunately, not every lexicon-based system is designed to obtain such high accuracies. In [Mirsa Karim and Smija Das, “Sentiment analysis on textual reviews,” *IOP Conference Series: Materials Science and Engineering*, vol. 396, p. 012020, 2018, doi: 10.1088/1757-899x/396/1/012020.] the researchers decided to compare the performance of SentiWord Net and Sentiment Vader with the performance of a naïve-bayes LDA. Based on the accuracy measurement obtained from the same data set, it is shown that SentiWord Net obtained an accuracy of 59.17% and Sentiment Vader obtained an accuracy of 54.7%, which are both considered low. Naïve-bayes obtained an accuracy of 75.2 which is much higher, compared to rule-based mechanisms. This actively demonstrates that machine learning techniques are possibly more accurate in calculating the sentiment of reviews.

# Machine Learning approaches

Various sentiment analysis systems have been built based on machine learning approaches. Hence, their investigation was deemed of paramount importance, from traditional machine learning to deep learning networks. The key difference between deep learning and traditional machine learning lies from the way data is presented to the system. Traditional machine learning algorithms almost always require structured data, whereas deep learning networks rely on layers of the ANN (artificial neural networks). Several studies compare and contrast the two approaches.

For example, in the work done in [G. Cai and B. Xia, “Convolutional Neural Networks for Multimedia Sentiment Analysis,” in *CCF International Conference on Natural Language Processing and Chinese Computing*, 2015, pp. 159–167.] the researchers chose to compare a deep network with a traditional network by using a CNN and Naïve Bayes, SVM and Logistic Regression for textual sentiment analysis. The proposed models were tested using two twitter datasets. The experimental results show that the proposed text CNN led to better performance than the other methods for text analysis.

Other studies also support the above, concluding that deep learning systems can offer a better performance than other traditional classifiers. This is proven by the work done in [C. Frans, Siti Mariyah, and Setia Pramana, “Sentiment analysis: a comparison of deep learning neural network algorithm with SVM and naϊve Bayes for Indonesian text,” *Journal of Physics: Conference Series*, vol. 971, p. 012049, 2018, doi: 10.1088/1742-6596/971/1/012049.] whereby the researchers compared the performance of a deep neural network with an SVM and a naïve-bayes classifier. The deep learning system manages to come out on top in all aspects. The model was tested with different feature extract techniques: Bag of Word, TF binary, TF-IDF and Bigram in order to have various results. In addition to this, unlike Naive Bayes and SVM, the modelling results produced by Deep Learning Neural Network are also not significantly affected by either balanced or unbalanced data composition conditions. On average, for the combination of Deep Learning Neural Network algorithms, the best feature extraction technique that can improve the final score of the modelling is Bigram (N-Grams). Thus, based on this research, one can conclude that the combination of Deep Learning Neural Network algorithm with bigram technique is very suitable for conducting sentiment analysis for Indonesian text data.

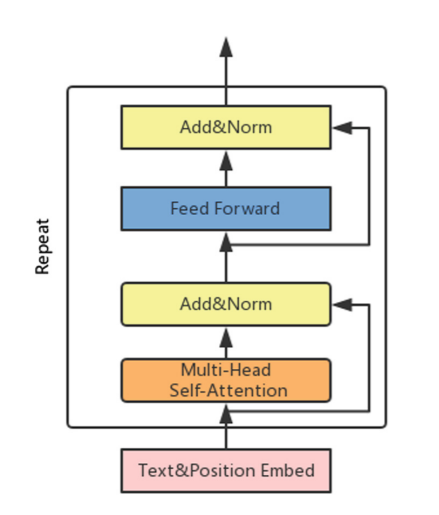
The work done in [K. Abu Kwaik, M. Saad, S. Chatzikyriakidis, and S. Dobnik, “LSTM-CNN Deep Learning Model for Sentiment Analysis of Dialectal Arabic,” *Communications in Computer and Information Science*, pp. 108–121, 2019, doi: 10.1007/978-3-030-32959-4\_8] also uses deep learning for classification. However, in this case, the researchers proposed a DL model that combines long-short term memory (LSTM) with convolutional neural networks (CNN). Here, a simple LSTM architecture was first introduced in order to determine its performance. This was tested on three dialectal SA datasets, which produced poor results. Then, an off-the-shelf SA model was taken from Kaggle, and its performance was tested. Such a model uses a combination of an LSTM and a CNN, and an improved performance was observed. Finally, the researchers proposed their own model, which is a more elaborate BiLSTM → CNN with more convolutional layers and obtained state-of-the-art results on the datasets that DL approaches have been previously applied to (i.e., the ASTD). In general, the results are promising but there is definitely room for improvement, especially on the three-way classification task. Of note, the researchers provided propositions which are believed to provide an ameliorated performance. These included the use of word embeddings specifically trained for the SA task, as well as even more complex DL architectures, for instance, ones using an attention mechanism.

Similarly, the work done in [A. Aslam, U. Qamar, P. Saqib, R. Ayesha, and A. Qadeer, “A Novel Framework for Sentiment Analysis Using Deep Learning,” in *International Conference on Advanced Communication Technology (ICACT))*, 2020, pp. 525–529, doi: 10.23919/ICACT48636.2020.9061247] involves a hybrid of different deep learning models namely Recurrent neural network(RNN), Long short Term Memory(LSTM), Bidirectional LSTM and Convolution neural network(CNN) and word2vec along with pre-trained word embedding. Each model is individually run-on input data and then the final prediction about polarity of text is produced, using majority voting on obtained results. Three datasets available online, namely SST-1, SST-2 and MR are used for experimental purposes. This proposed system managed to obtain an accuracy of 87.5% which is a very good accuracy when compared to other works.

S. Xiao*et al* also agree on the superiority of a deep learning system. [S. Xiao*et al*, "Sentiment Analysis for Product Reviews Based on Deep Learning," *Journal of Physics: Conference Series,*vol. 1651, *(1),*2020. Available: https://ejournals.um.edu.mt/login?url=https://www-proquest-com.ejournals.um.edu.mt/scholarly-journals/sentiment-analysis-product-reviews-based-on-deep/docview/2513091833/se-2?accountid=27934. DOI: <http://dx.doi.org.ejournals.um.edu.mt/10.1088/1742-6596/1651/1/012103>.] In their work, they compared the performance of an LSTM algorithm with the performance of a naïve-bayes and logistic regression model. In contrast to other research, such as that of [C. Frans, Siti Mariyah, and Setia Pramana, “Sentiment analysis: a comparison of deep learning neural network algorithm with SVM and naϊve Bayes for Indonesian text,” *Journal of Physics: Conference Series*, vol. 971, p. 012049, 2018, doi: 10.1088/1742-6596/971/1/012049.], Shaozhang Xiao et al. opted for a different feature extraction. The feature extraction method was the word2vec, whereby the sample data is converted into a vector matrix. The accuracy of the three algorithms, namely, Naive Bayes, logistic regression, and LSTM in the first dataset reaches about 80%. However, the accuracy rate of the models in the data set related to Weibo reviews is only about 60%, indicating that the fitting ability of the models in the unseen data set is not adequate. This is attributed to the lack of corresponding features of training samples. Since this system is mainly applied to product reviews, the training effect of the models has reached expectation. LSTM exhibits good performance in product analysis in the system. It also performs well in predicting product reviews, but it does not achieve satisfying results in the prediction of some short texts or unseen data set, hence necessitating further improvement and optimization. Moreover, the data sets of this system are mostly data sets on product reviews. The model has almost learned all the features of the data set after being trained to a certain extent, thus the effect of continuing to train the model is not good. This can be improved by expanding the range of the data set. Nevertheless, in all cases, the LSTM proved to have an edge over the other algorithms, also concluding that deep learning algorithms give more accurate results in SA.

Although in the work discussed above, deep learning networks succeeded to outperform traditional networks, the following study concluded that this was not the case. In the work done by Nichole et al. [Nichole thesis], the classifiers used within the CFC (Custom Feature Component), requiring hand-crafted features, outperformed the deep learning network used within the UFC (Unsupervised Feature Component). This allowed the researchers to conclude that the former classifiers were more suitable for classification of Maltese text within their research rather than the latter classifiers. Hence, although it is still unclear as to which machine learning approach is superior to the other, one can safely say that the two presents with their own advantages and disadvantages respectively.

The work done in [Y. Pan, B. Song, N. Luo, X. Chen, and H. Cui, “Transformer and multi-scale Convolution for target-oriented Sentiment Analysis,” in *Asia-Pacific Web (APWeb) and Web-Age Information Management (WAIM) Joint International Conference on Web and Big Data*, 2019, vol. 11642, pp. 310–321.] involves the use of transformers to solve the problem of sentiment analysis. The researchers have designed a transformer of the following structure:



The model’s performance was compared to several other machine learning systems. The designed model produced the best performance on the three datasets, having 84.20% accuracy on restaurant dataset, 78.21% accuracy on laptop dataset, and 72.98% accuracy on twitter dataset. LSTM has the worst performance of all neural networks. ATAE-LSTM improves its performance by taking the target into account and using the attention mechanism. IAN works better because it uses two attention layers. For the twitter dataset, BILSTM ATT-G and RAM cannot perform as efficiently as they do in restaurant and laptop dataset, because they are heavily rooted in LSTM, which is not ideal at processing ungrammatical sentences. TNet is a model based on LSTM and CNN, which makes it works well on all the three datasets. Different from previous models, the model produced by the researchers is based on transformers, which can solve long term dependencies and can be easily parallelized. The multi-scale convolution layer in their model can extract multi-grained features. The transformer and multi-scale convolution structure help their model get the best performance on all the three datasets.

Similarly, a transformer-based system was designed in [Y. Xie, H. Wen and Q. Yang, "Ternary Sentiment Classification of Airline Passengers’ Twitter Text Based on BERT," *Journal of Physics: Conference Series,*vol. 1813, *(1),*2021. Available: https://ejournals.um.edu.mt/login?url=https://www-proquest-com.ejournals.um.edu.mt/scholarly-journals/ternary-sentiment-classification-airline/docview/2512976750/se-2?accountid=27934. DOI: http://dx.doi.org.ejournals.um.edu.mt/10.1088/1742-6596/1813/1/012017.] to tackle the sentiment analysis problem. the ternary sentiment classification model is based on BERT (Bidirectional Encoder Representations from Transformers). The comparative experiments indicate that the proposed model is advantage over the model based on ELMo, and the evaluation indexes verify, the excellent performance of BERT in the field of word representation. The work of this paper has a certain significance for the research of UGC (user generated content) text sentiment classification. UGC text has the characteristics of concise words and networked expressions, words of UGC text have rich connotations. By using BERT, the semantic feature vector of the UGC text can be obtained, and the polysemy problem in different contexts can be solved, which provides new solutions and research ideas for UGC text sentiment classification. However, this method also has some disadvantages, such as running BERT on the computer has high computational complexity.

# Hybrid solutions

Other than the traditional and deep approaches, some researchers opt to make use of hybrid solutions, which involves combining the use of machine learning with another technique. Their aim is to construct a sentiment analysis system which achieves better results.

For instance, in [ElBeltagy, Samhaa R, T. Khalil, A. Halaby, and M. Hammad, “Combining Lexical Features and a Supervised Learning Approach for Arabic Sentiment Analysis,” in *Computational Linguistics and Intelligent Text Processing*, 2018, pp. 307–319.] the researchers have presented a system that uses a lexicon to match the input data against entries in a sentiment lexicon. The researchers have also presented a set of features that can be used with Arabic tweets in order to enhance the performance of a sentiment analysis system. ElBeltagy et al believe that features derived from a high-quality sentiment lexicon provide the highest impact on enhancing the results. By using these features within a sentiment analysis system, they devise a model that outperforms all existing sentiment analysis systems on 6 out of the 7 datasets on which they have applied it. The models used are a Complement Naïve Bayes (CNB) classifier, an SVM and a multi-nominal updateable Naïve Bayes Classifier. The results demonstrate that even when using the default configuration (which is not always the best), the presented model outperforms all other existing systems on all datasets except for RR2.

Similarly, O. Appel et al [O. Appel, F. Chiclana, J. Carter, and H. Fujita, “A Hybrid Approach to Sentiment Analysis with Benchmarking Results,” in *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, 2016, pp. 242–254.] present a hybrid approach to Sentiment Analysis encompassing the use of Semantic Rules, Fuzzy Sets and an enriched Sentiment Lexicon, improved with the support of SentiWordNet. The proposed hybrid method is compared against two well-established Supervised Learning techniques, Naive Bayes, and Maximum Entropy. Using the well-known and publicly available Movie Review Dataset, the proposed hybrid system achieved higher accuracy and precision than Naive Bayes (NB) and Maximum Entropy (ME). In general, the proposed hybrid system works very well with a high level of accuracy and precision. In this work the use of efficient NLP techniques (like tokenizing, parsing, negation handling, etc.), contributed positively to the application of a Hybrid Method. Moreover, the creation of an improved Sentiment Lexicon was a critical component in obtaining good experimental results. Additionally, SentiWordNet was established as a crucial element of the proposed solution and certainly enriched the quality of the Lexicon dramatically.

As a continuation to this, another hybrid system making use of SentiWordNet was studied. The work done in [J. Zhang, "A Combination of Lexicon-Based and Classified-Based methods for Sentiment Classification based on Bert," *Journal of Physics: Conference Series,*vol. 1802, *(3),*2021. Available: https://ejournals.um.edu.mt/login?url=https://www-proquest-com.ejournals.um.edu.mt/scholarly-journals/combination-lexicon-based-classified-methods/docview/2512965771/se-2?accountid=27934. DOI: http://dx.doi.org.ejournals.um.edu.mt/10.1088/1742-6596/1802/3/032113.] combines the use of SentiWordNet with a transformer-based system. The model was compared with two state-of-the-art methods. The proposed model is based on BERT-BASE, were the researchers used the existing pre-trained model, and just fine-tuned the model according to their dataset. their method beats the other two methods and outperforms the SO-CAL significantly. The researchers’ model used < 𝑠𝑒𝑝 > tags to collect the sentiment information from each clause and combine them through the < 𝑐𝑙𝑠 > tag to make the prediction, which made the model more suitable for the long sequence with many sentiment features. The designed model not only has much faster speed than the BERT-BASE model in training and inference speeds but also obtained a higher F-1 score than the BERT-BASE.

# List of approaches

|  |  |  |
| --- | --- | --- |
| ML Approach – Maltese | CFC and UFC | 62.3% |
| ML Approach - Maltese | SVM and RF | Very low |
| Lexicon-Based | No of positive vs no of negative in sentence | 68% |
| Lexicon-Based | Word vector. Unreliability score |  |
| Lexicon-Based | AffinPDO | 81% |
| Lexicon-Based | SentiwordNet, Sentiment Vader | 59%, 54% |
| ML Approach | CNN, Naïve Bayes, SVM and Logistic Regression | 0.74, 0.77 – 0.70, 0.72 – 0.72, 0.74 – 0.73, 0.76 |
| ML Approach | Naïve Bayes, SVM, Deep Neural Net | 0.85, 0.88, 0.99 |
| ML Approach | Deep Neural Net | 76%, 66%, 68% |
| ML Approach | Hybrid of different deep learning models namely RNN, LSTM, Bidirectional LSTM and  CNN | 87.5%  (Binary classification) |
| ML Approach | Naive Bayes, logistic regression, and LSTM | 85.07%, 86.05%, 89.85% |
| ML Approach | Transformer | 78% |
| ML Approach | Transformer - BERT | 78.66% |
| Hybrid Approach |  | 81 (MNBU) 85.03 (SVM) 71.06 (CNB) 80.6 (SVM) 83.13 (MNBU) |
| Hybrid Approach | SentiWordNet + NaiveBayes and Maximum Entropy | 0.76 |
| Hybrid Approach | SentiWordNet + Transformer | 0.83 |