2023 14th International Conference on Intelligent Systems: Theories and Applications (SITA) | 979-8-3503-0821-1/23/$31.00 ©2023 IEEE | DOI: 10.1109/SITA60746.2023.10373715

# Call Center Customer Sentiment Analysis Using ML and NLP

1st Imad AATTOURI
*Ain Chock Faculty of Science*
*Hassan 2 University*
Casablanca, Morocco
aattouriimad@gmail.com

2nd Hicham MOUNCIF
*Polydisciplinary Faculty of Beni Mellal*
*Sultan Moulay Slimane University*
Beni Mellal, Morocco
h.mouncif@usms.ma

3rd Mohamed RIDA
*Ain Chock Faculty of Science*
*Hassan 2 University*
Casablanca, Morocco
mhd.rida@gmail.com

***Abstract*—In the contemporary digital era, call centers have significantly incorporated automation through callbots, but they often lack an essential aspect of customer service: empathy. This paper explores the integration of sentiment analysis into call center operations to introduce an emotional dimension to callbot interactions. Utilizing natural language processing and machine learning, the paper examines both text-based and signal-based sentiment analysis approaches. The proposed sentiment analysis architecture encompasses data input and output interfaces, preprocessing, sentiment analysis, and a decision manager module that can, for example, escalate calls to a human agent based on sentiment analysis results. The preprocessing steps, both for text and signal analysis, are outlined in detail, along with a review of relevant sentiment analysis algorithms.**

**Through comprehensive experiments, the study demonstrates that the integration of sentiment analysis yields promising outcomes. In text-based analysis, SVM and LSTM models consistently performed well, achieving accuracy scores of 74% and 72%, respectively. For voice-based analysis, the MLP model exhibited the highest accuracy of 0.72 when using mel spectrogram features, while the RF model outperformed others with an accuracy of 0.78 using MFCC features. These results showcase the potential of sentiment analysis in humanizing callbot interactions, thereby enhancing customer satisfaction and service efficiency.**

***Index Terms*—Sentiment Analysis, Call Centers, Callbots, NLP , Machine Learning, Text-based Analysis, Signal-based Analysis, Emotion Detection, SVM, Naive Bayes, Random Forest, LSTM, MLP, Mel Frequency Cepstral Coefficients (MFCC), Customer Satisfaction.**

## I. Introduction

In the era of digitalization and artificial intelligence, customer interaction has undergone a significant transformation. Nowadays, it is common to reach out to customer service and be greeted not by a human voice, but by a callbot. These automated systems, powered by complex algorithms and sometimes even neural networks, aim to streamline and expedite the handling of customer requests, while optimizing human resources. However, despite their technical efficiency, these callbots can lack one essential thing: empathy.

Feeling and emotion play a pivotal role in human communication. A displeased, frustrated, or confused customer requires a different approach than a satisfied or neutral one. In a traditional call center, a human agent can discern these nuances through the customer's tone, choice of words, and voice inflections. But a standard callbot can't do this... at least, not without the right technological aid.

This is where sentiment analysis, combining ML techniques and NLP , comes into play. By equipping callbots with this capability, it becomes possible to detect a customer's sentiment and act accordingly, whether it's to adjust the bot's response or to transfer the call to a human agent.

In this article, we will explore how such integration could not only humanize automated interactions but also greatly enhance the customer experience in call centers. We will outline the techniques and methodologies we've adopted, focusing on two main approaches: text-based analysis and signal-based vocal analysis.

979-8-3503-0821-1/23/$31.00 ©2023 IEEE

## II. Literature Review

This article is a continuation of our work on the development of an intelligent callbot [1], [2], [3].

Sentiment analysis is a firmly established subfield of NLP that endeavors to ascertain the emotion or sentiment conveyed within a textual entity [4]. It has found manifold applications, from social media analysis [5] to finance [6], and, of course, call centers.

### A. Text-based Sentiment Analysis in Call Centers

Text-based sentiment analysis techniques have been extensively studied and applied to various domains. Within the call center context, call transcriptions can be analyzed to gauge customer sentiment [7]. Common methods include those based on bag-of-words models, such as SVM [8] and logistic regression [9], as well as more advanced techniques employing neural networks like LSTM [10] and BERT [11].

### B. Signal-based Sentiment Analysis

Beyond lexical content, voice tone, pacing, pauses, and other prosodic features can carry significant insights into a speaker's sentiment or emotional state [12]. Techniques like Convolutional Neural Networks (CNN) [13] and Gaussian Mixture Models (GMM) [14] have been employed to classify emotions from voice signals.

### C. Challenges in Emotion Detection for Call Centers

Accurate emotion detection within call centers presents unique challenges. In addition to background noise that may hamper signal quality [15], individual variability in emotion expression [16], and the need for real-time data processing [17] complicate the endeavor.

## III. Materials and Methods

### A. Materials

In this article, we explore the integration of sentiment analysis into call center operations to introduce an emotional dimension into interactions with callbots. To capture this emotional dimension from two complementary perspectives, we utilize two distinct datasets: one based on voice recordings and another based on text transcriptions. This methodology allows for a more comprehensive comprehension of the sentiments conveyed by customers during their interactions with callbots.

The use of two different datasets is justified for several reasons. First, vocal and text-based interactions differ in how emotions are conveyed. Emotions in the voice can be transmitted through elements such as tone, intensity, and pace, whereas in text, emotions rely on word choices, expressions, and phrasing. By combining voice recordings with text transcriptions, we encompass a broader range of emotional cues.

Furthermore, the combination of the two datasets enhances the robustness and generalizability of our sentiment analysis approach. ML models trained on multiple datasets better capture variations and nuances in the sentiments expressed by customers.

*1) Text-based Sentiment DataSet:* For the text-based sentiment analysis aspect of our research, the selection of an appropriate dataset played a pivotal role. After careful consideration of various available options, we opted for the "French Twitter Sentiment Analysis" dataset [18], which aligns closely with the objectives of our study. With 1.5 million tweets, the dataset offers a substantial amount of textual content for analysis, enabling us to obtain robust and comprehensive insights into sentiment patterns.

*2) voice-based Sentiment DataSet:* For our voice-based sentiment analysis, we harnessed the "Canadian French Emotional (CaFE) Speech Dataset" [19], a notable resource in the field. This dataset features a diverse array of emotional speech recordings in the French language, comprising six distinct sentences uttered by twelve actors representing both genders. Covering six fundamental emotions and a neutral state, the CaFE dataset consists of three parts. Initially, the dataset encompassed 864 audio files capturing various emotional expressions.

### B. Methods

*1) Sentiment Analysis Architecture Proposed for the Call Center:* In order to facilitate a comprehensive analysis of customer interactions within the call center, we have designed a sentiment analysis architecture as depicted in Figure 1. This architecture provides an overview of the consecutive steps involved in assessing and responding to the feelings of customers as they interact with a callbot.

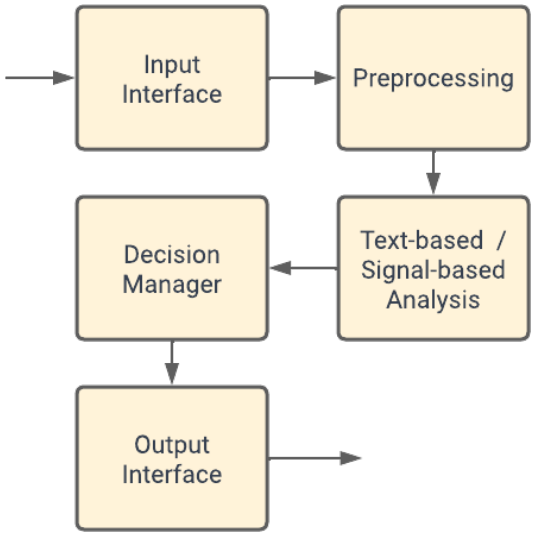


Fig. 1. Sentiment Analysis Architecture for the Call Center.

*a) Input Interface:* The first step in this architecture captures the client's interaction. Whether it's a vocal or textual communication, this interface is pivotal as it stands as the entry point for all the data the system needs to analyze.

*b) Preprocessing:* After capturing, the data undergoes a preprocessing phase. In this step, any superfluous or irrelevant information is filtered out, and the data is formatted for optimized analysis. This phase ensures only pertinent data is forwarded to the analysis stage.

*c) Analysis:* This is the heart of the architecture. Once preprocessed, the data is analyzed to discern the client's sentiment. Employing advanced ML and NLP techniques, the system can evaluate whether the client is satisfied, frustrated, neutral, or exhibits any other relevant emotion.

*d) Decision Manager:* Based on the results from the analysis phase, the decision manager determines the next course of action. If, for instance, the client seems discontented or frustrated, the manager might decide to transfer the call to a human agent for a more personalized handling.

*e) Output Interface:* Lastly, this final step implements the determined decision. This could manifest as an automated response to the client, a closing message, or a transfer to a human agent, depending on what the decision manager has determined.

*2) Preprocessing for Text-based Analysis:* To ensure accurate sentiment analysis on text data, a series of preprocessing steps (Figure 2) were applied to the extracted content from the dataset.

*a) Text Cleaning:* The text data underwent thorough cleaning, involving the removal of punctuation, special char-

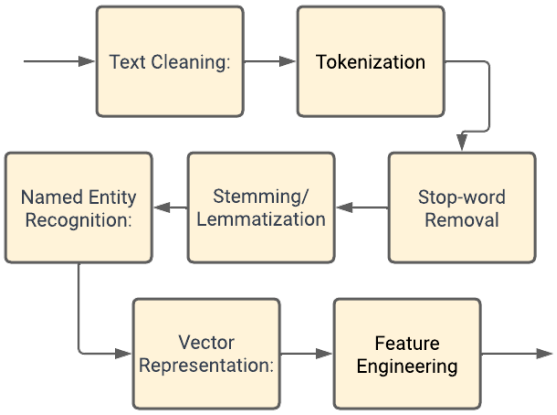


Fig. 2. Preprocessing for Text-based Analysis

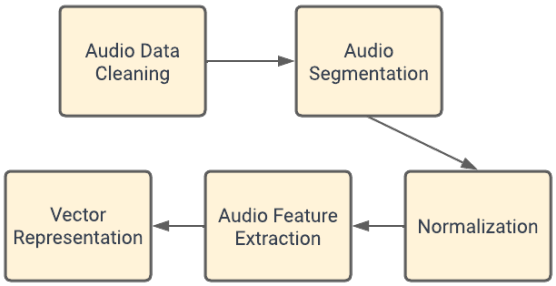


Fig. 3. Preprocessing for Voice-based Analysis

acters, and numerical digits. This step aimed to reduce noise and standardize the data.

*b) Tokenization:* Tokenization segmented the text into individual words or tokens. This process provided the basis for subsequent analysis, breaking down the text into manageable units.

*c) Stop-word Removal:* Common stop words, such as articles, conjunctions, and prepositions, were eliminated from the text to focus on significant content and enhance the quality of analysis.

*d) Stemming/Lemmatization:* Stemming or lemmatization was applied to reduce words to their base forms. This step aimed to normalize variations and facilitate more meaningful analysis.

*e) Named Entity Recognition:* Named Entity Recognition (NER) identified and categorized proper nouns, such as names of people, places, and organizations. NER enriched the understanding of context and entities within the text.

*f) Vector Representation:* Text data was transformed into numerical vectors through techniques like CountVectorizer and TF-IDF. This conversion enabled ML algorithms to process and analyze the text.

*g) Feature Engineering:* Additional features were engineered to enhance analysis, including n-grams to capture contextual information and phrase structures.

The comprehensive preprocessing process lays the foundation for accurate sentiment analysis, enabling the extraction of insightful emotional expressions from the call center context.

*3) Preprocessing for Voice-Based Analysis:* In order to facilitate a comprehensive sentiment analysis on voice data, a series of preprocessing steps were applied to the "Canadian French Emotional (CaFE) Speech Dataset." (Figure 3).

*a) Audio Data Cleaning:* The initial preprocessing step involved thorough cleaning of the audio recordings within the CaFE dataset. Background noise, artifacts, and potential distortions were meticulously removed to enhance the clarity and fidelity of the voice data.

*b) Audio Segmentation:* Continuous audio streams from the CaFE dataset were segmented into smaller, meaningful units, such as sentences or utterances. This segmentation process not only organized the data but also allowed for focused analysis of distinct emotional expressions.

*c) Audio Feature Extraction:* Relevant acoustic features were extracted from each segmented audio unit. These features included pitch, intensity, speech rate, and pauses, among others. The extraction process transformed raw audio data into quantifiable attributes for subsequent analysis.

*d) Normalization:* To ensure consistency and comparability across various recordings, the extracted audio features underwent normalization. This step standardized the range and scale of features, facilitating accurate and meaningful analysis.

*e) Vector Representation:* Normalized audio features were then converted into numerical vectors, rendering them suitable for input to ML algorithms. This vector representation formed the basis for our sentiment analysis techniques.

The preprocessing of voice data from the Canadian French Emotional (CaFE) Speech Dataset lays a solid groundwork for our subsequent sentiment analysis, enabling us to delve into the intricate emotional nuances present within call center interactions.

*4) Algorithms:*

*a) SVM (Support Vector Machine):* is a classification algorithm used in sentiment analysis to separate positive and negative comments based on textual features. It seeks the optimal hyperplane that maximizes the margin between classes, effectively classifying new examples based on their proximity to this hyperplane[20].

*b) Naive Bayes:* is an algorithm that employs the principles of Bayes' theorem in order to forecast the category of a given sample. This is achieved by calculating the conditional probability of each class, taking into account the input features. Despite its simplistic assumption of independence among features, NB can be effective for sentiment analysis tasks.[21]

*c) Random Forest:* is a collective algorithm that constructs numerous decision trees in the process of training and merges their forecasts to generate the ultimate outcome. In the context of sentiment analysis, Random Forest has the capability to apprehend intricate associations among character-

istics and sentiments, consequently augmenting the precision of predictions.[22]

*d) Long Short-Term Memory:* LSTM is an Recurrent Neural Network(RNN) created to solve the vanishing gradient issue in sequence modeling. It can capture long-term dependencies in sequential data, making it useful for sentiment analysis tasks involving text sequences..[23]

*e) Multi-Layer Perceptron:* a category of artificial neural networks is recognized for its numerous layers of interconnected nodes. It encompasses an initial layer for input, one or more concealed layers, and a concluding output layer. The training of MLPs is achieved through the utilization of activation functions and backpropagation, rendering them adaptable for a multitude of tasks, such as sentiment analysis.[24]

*5) Text Vector Representation Techniques:* Leveraging appropriate text vector representation techniques is crucial for effective sentiment analysis and other NLP tasks. In this section, we delve into two widely used techniques: TF-IDF and CountVectorizer.

*a) TF-IDF (Term Frequency-Inverse Document Frequency):* TF-IDF is a fundamental text vectorization method that evaluates the importance of words in a document relative to their frequency across a corpus. The technique combines Term Frequency (TF), which measures the occurrence of a word within a specific document, and Inverse Document Frequency (IDF), which gauges the rarity of the word across the entire collection. [25]

*b) CountVectorizer:* This method is a straightforward and efficient approach for transforming textual data into numerical vectors. Its functioning involves the creation of a matrix, where each row represents a document and each column represents a distinct word from the complete collection of texts. The entries in the matrix indicate the frequency at which each word appears in its respective document.[26]

Both TF-IDF and CountVectorizer are valuable tools for transforming raw text into structured numerical data, enabling ML algorithms to effectively analyze and classify text documents for sentiment analysis and various other applications.

*6) Audio Feature Extraction:* Efficient audio feature extraction is crucial for converting raw audio signals into meaningful numerical representations that can be utilized for sentiment analysis and other audio processing tasks. In this section, we delve into two prominent techniques: the Mel Spectrogram and MFCCs.

*a) Mel Spectrogram:* The Mel Spectrogram is a widely utilized technique in the field of audio analysis, which provides a visual representation of the frequency characteristics of an audio signal over a period of time. Its effectiveness in capturing variations in pitch and timbre makes it extremely valuable for detecting subtle emotional nuances in speech. The Mel Spectrogram divides the audio signal into short overlapping intervals, calculates the magnitude of the Fourier Transform for each interval, and then maps the resulting spectrum onto the Mel scale—a frequency scale that is perceptually significant. This process generates a two-dimensional matrix where one axis represents time, another represents frequency, and the intensity of each matrix element indicates the magnitude of the corresponding frequency component. The Mel Spectrogram serves as a fundamental feature for various audio-related tasks, such as analyzing the sentiment in call center interactions.[27]

*b) MFCCs (Mel Frequency Cepstral Coefficients):* MFCCs represent a succinct portrayal of an audio signal's spectral characteristics. Commonly employed in speech and audio processing, they capture the unique timbral and phonetic attributes of a sound. The MFCC extraction process involves multiple steps, encompassing computation of the Mel Spectrogram, followed by the application of Discrete Cosine Transform (DCT) to decorrelate the Mel frequency elements. The resultant coefficients outline the audio signal's spectral envelope shape, effectively compressing information while retaining pertinent attributes. MFCCs find extensive use in automatic speech recognition and emotion recognition tasks, rendering them a valuable tool for extracting distinguishing features from call center voice recordings.[27]

Both the Mel Spectrogram and MFCCs play a pivotal role in metamorphosing raw audio data into meaningful features that encapsulate the auditory traits of speech. These features empower sentiment analysis models to decode emotional expressions and sentiments conveyed through voice interactions in the call center milieu.

*7) Comparison Metrics:* To assess the performance of the sentiment analysis models applied to call center interactions, several key metrics are employed:

*a) Recall:* Recall measures the model's ability to correctly identify actual positive instances, calculated as:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

*b) F1-Score:* is the harmonic mean of precision and recall, offers a balanced assessment of a model's performance. This metric takes into account both false positives and false negatives and is particularly valuable when working with imbalanced datasets. It is computed using the following formula.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

*c) Precision:* The precision metric determines the correctness of positive predictions by calculating the ratio of accurately predicted positive instances to all instances predicted as positive, which can be computed using the provided formula.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

*d) Accuracy:* The accuracy of a model's predictions is determined by the proportion of correctly classified instances out of the total number of instances.

$$\text{Accuracy} = \frac{\text{True Positives} + \text{True Negatives}}{\text{Total Instances}}$$

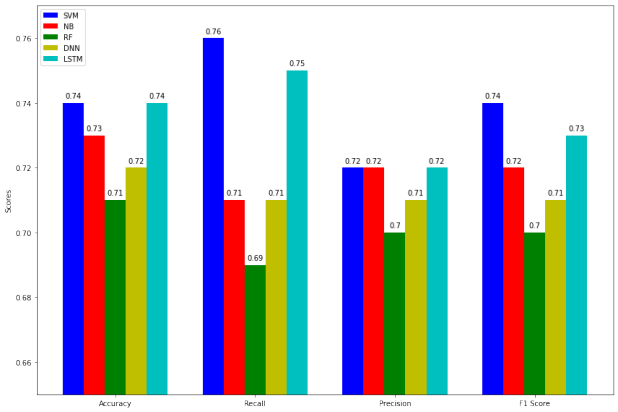


Fig. 4. Performance of Text Analysis Algorithms with CountVectorizer.

## IV. Results and Discussion

### A. Text-based Analysis

When utilizing the CountVectorizer for feature extraction, the models demonstrated comparable performances. Specifically, both the SVM and LSTM algorithms displayed the highest accuracy, achieving a score of 0.74. The recall metric, which measures the model's ability to correctly identify all relevant instances, showed the LSTM slightly ahead with a score of 0.75. When evaluating precision, which represents the accuracy of the positive predictions, all the models gravitated around the 0.70-0.72 range. The F1 scores, which balance precision and recall, further accentuated the superior performance of the SVM and LSTM models.

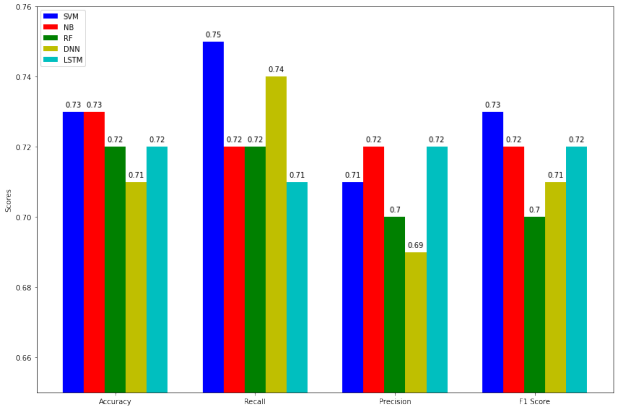


Fig. 5. Performance of Text Analysis Algorithms with TfidfVectorizer.

For the TfidfVectorizer, the performances slightly dwindled in comparison to the CountVectorizer results. The SVM achieved the pinnacle of accuracy with a score of 0.73. An interesting observation arose in the recall department, with the MLP model showcasing the highest value at 0.74. Precision scores remained consistent with the previous set, floating within the 0.69-0.72 spectrum. Finally, the F1 scores reemphasized the balanced and steady performances of all models.

### B. voice-based Analysis

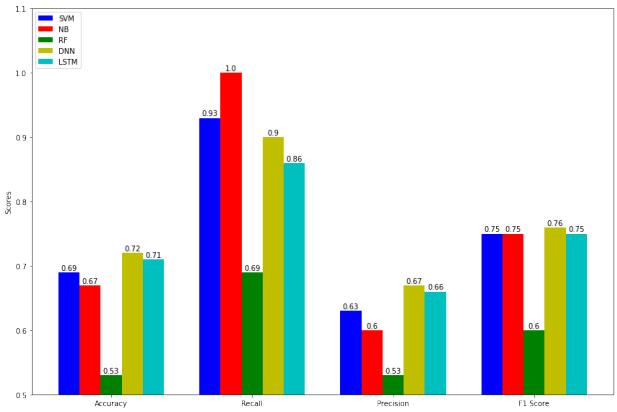


Fig. 6. Performance of voice Analysis Algorithms with melspectrogram.

Diving into the realm of voice-based analysis using the melspectrogram, the MLP model emerged as the frontrunner with an accuracy score of 0.72. An intriguing facet was the NB model's impeccable recall of 1.00. While this might seem impressive at first glance, it warrants careful scrutiny as it might insinuate the model's propensity to overfit or possess a bias towards the positive class. Precision values, which were mainly sandwiched between 0.60 and 0.67, didn't showcase any standout performer. However, in terms of F1 scores, the MLP and LSTM shined once more.

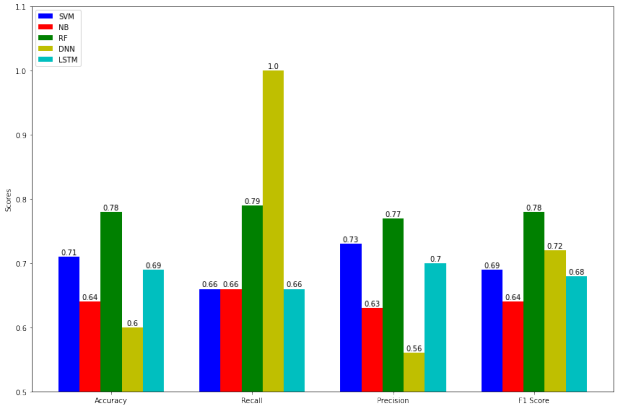


Fig. 7. Performance of voice Analysis Algorithms with mfccs.

When the voice-based features were represented using MFCCs, a significant shift in model performance was evident. The RF algorithm notably outpaced its competitors with an impressive accuracy of 0.78. Mirroring the previous set's trend, the MLP boasted a flawless recall score of 1.00. However, as mentioned earlier, such perfect scores require cautious interpretation. The RF maintained its supremacy in precision with a score of 0.77, and similarly, the F1 scores portrayed it as the best performer.

In conclusion, for text-based analysis, while there were slight discrepancies between the two vectorization techniques, SVM and LSTM consistently emerged as top contenders. On

the contrary, in voice-based analysis, feature representation played a pivotal role in discerning the best models: MLP and LSTM for mel spectrogram and RF for MFCCs. However, it's paramount to consider factors such as interpretability, efficiency, and real-world application needs before finalizing any model for deployment. Moreover, while standard metrics provide a robust understanding of model performance, an in-depth examination, especially regarding edge cases, remains indispensable.

## V. Conclusion

The integration of sentiment analysis in call centers offers a promising opportunity to enhance the customer experience by introducing an emotional dimension into automated interactions. Through the use of ML and NLP techniques, it becomes possible to detect customer sentiments from text and voice signals, paving the way for tailored responses and intelligent call routing. The experiments have shown promising results, but there are challenges to overcome, including adaptation to different accents and languages, as well as ongoing improvement of sentiment analysis models. The future of automated call centers could be much more than mechanical interaction. With the addition of artificial empathy, these systems could become true partners in delivering exceptional customer service.

In future endeavors, we plan to advance our research by focusing on the seamless integration of the sentiment analysis module within the callbot architecture. This integration aims to provide real-time monitoring and analysis of customer interactions, offering valuable insights into the callbot's performance and the emotional dynamics of the conversations.

## References


[1] I. Aattouri, M. Rida, and H. Mouncif, "A comparative study of learning algorithms on a call flow entering of a call center," 2021. DOI: 10.1007/978-3-030-73103-8_36.

[2] I. Aattouri, M. Rida, and H. Mouncif, "Creation of a callbot module for automatic processing of a customer service calls," 2021. DOI: 10.1007/978-3-030-76508-8_30.

[3] I. Aattouri, H. Mouncif, and M. Rida, "Modeling of an artificial intelligence based enterprise callbot with natural language processing and machine learning algorithms," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 12, pp. 943–955, 2 2023. DOI: 10.11591/ijai.v12.i2.pp943-955.

[4] B. Liu, *Sentiment analysis and opinion mining* (Synthesis lectures on human language technologies 1). 2012, vol. 5, pp. 1–167.

[5] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," in *LREC*, 2010.

[6] B. Wuthrich *et al.*, "Daily stock market forecast from textual web data," in *IEEE International Conference on Systems, Man, and Cybernetics*, 1998.

[7] L. Zhao *et al.*, "Call center customer complaint prediction with topic modeling and sentiment analysis," in *International Conference on Web Information Systems Engineering*, 2012.

[8] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[9] D. Jurafsky and J. H. Martin, *Speech and language processing*. 2019.

[10] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[11] J. Devlin *et al.*, "Bert: Pre-training of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[12] B. Schuller *et al.*, "A tutorial on paralinguistics in speech and language processing," *Proceedings of the IEEE*, 2011.

[13] Z. Zhang *et al.*, "Pattern recognition in speech and language processing," *Electrical Engineering & Automation*, 2017.

[14] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models," *IEEE transactions on speech and audio processing*, 1995.

[15] U. Zölzer, Ed., *DAFX: Digital Audio Effects*. John Wiley & Sons, 2011.

[16] M. El Ayadi, M. S. Kamel, and F. Karray, "Survey on speech emotion recognition: Features, classification schemes, and databases," *Pattern Recognition*, vol. 44, no. 3, pp. 572–587, 2011.

[17] E. Ribeiro *et al.*, "Real-time speech emotion recognition using gpu and its applicability in neurology," *Health Information Science and Systems*, vol. 3, no. S1, 2015.

[18] "French twitter sentiment analysis dataset." (), [Online]. Available: https://www.kaggle.com/datasets/hbaflast/french-twitter-sentiment-analysis.

[19] "Canadian french emotional (cafe) speech dataset." (), [Online]. Available: https://zenodo.org/record/1478765.

[20] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[21] A. Y. Ng and M. I. Jordan, "The optimality of naive bayes," *Advances in neural information processing systems*, vol. 14, pp. 849–856, 2001.

[22] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[23] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[24] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, no. 6088, pp. 533–536, 1986.

[25] K. S. Jones and S. Walker, "A study of automatic text classification," *Journal of Documentation*, vol. 53, no. 1, pp. 1–32, 1997.

[26] C. D. Manning and H. Schütze, *Foundations of Statistical Natural Language Processing*. The MIT Press, 1999.

[27] J. P. Bello and L. R. Rabiner, "Mel frequency cepstral coefficients for music modeling," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2010, pp. 5542–5545.