Leveraging AI for Multidimensional Sentiment Analysis to Automate Customer Feedback Response within Salesforce CRM

MSc Research Project
Artificial Intelligence for Business

Barbaros Sonmez
Student ID: x23169788

School of Computing
National College of Ireland

Supervisor: Faithful Onwuegbuche

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Barbaros Sonmez
Student ID:	x23169788
Programme:	Artificial Intelligence for Business
Year:	2024
Module:	MSc Research Project
Supervisor:	Faithful Onwuegbuche
Submission Due Date:	12/08/2024
Project Title:	Leveraging AI for Multidimensional Sentiment Analysis to Automate Customer Feedback Response within Salesforce CRM
Word Count:	6021
Page Count:	20

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature:	
Date:	15th September 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST:

Attach a completed copy of this sheet to each project (including multiple copies).	☐
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	☐
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	☐

Assignments that are submitted to the Programme Coordinator office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Leveraging AI for Multidimensional Sentiment Analysis to Automate Customer Feedback Response within Salesforce CRM

Barbaros Sonmez
x23169788

Abstract

Today's customer-centric industry has forced businesses of all sizes to gain insight into their customers. Sentiment analysis is an important tool for analyzing large amounts of consumer feedback and understanding customers. Classic sentiment analysis categorizes utterances as positive or negative, however multi-dimensional sentiment analysis provides more information about language, such as emotions. Salesforce, as a Customer-Relationship-Management (CRM) solution, includes consumer feedback in the form of reviews, emails, and comments, which could offer organizations with useful information to improve their service and products. The literature review shows that there are not enough studies on using machine learning to improve the capabilities of CRM systems. Improving Salesforce's machine learning capabilities for multi-dimensional sentiment analysis on customer feedback data can lead to increased customer understanding. This study aims to answer the question of how multi-dimensional sentiment analysis could leverage customer feedback to create an automated response system within Salesforce CRM. For this purpose, Support Vector Machine (SVM), RoBERTa, and Electra models were employed to perform multi-dimensional sentiment analysis on the GoEmotions dataset. The best performing algorithm Electra was connected to Salesforce to establish a tailored and customized customer response. The implementation resulted in accurate multi-dimensional sentiment analysis and timely response to customers within Salesforce, significantly improving CRM capabilities.

1 Introduction

All organizations must comprehend how their customers feel and prioritize improving their experience. The sheer volume of data to be processed necessitates the use of automated methods to succeed in today's ever-changing markets (Gowri et al., 2022). Customer feedback generates useful information to improve the products and services. Understanding and interpreting this input is critical. The output of customer feedback analysis is used to define corporate goals, therefore playing a key role in evaluating the outputs and determining the best strategy for action (Charitha et al., 2023). Particularly, for large organizations with a large number of consumers, it is vital that these operations are performed instantly utilizing machine learning (ML) (Dieksona et al., 2023). As demonstrated in the section on related work, there is not enough research that investigates how the capabilities of Customer-Relationship-Management (CRM) could be utilized to handle customer interactions and understand customers through the application of ML algorithms. This research aims to add numerical evaluation scores to the existing literature on this topic.

Salesforce CRM is a cloud-based customer relationship management software. It was the market leader for ten consecutive years. The market statistic is shown in Figure 1.

Figure 1: Market share of popular CRM software providers (see `https://www.salesforce.com/news/stories/idc-crm-market-share-ranking-2023/`).

Salesforce CRM holds a huge amount of information including customer feedback. This information is stored in the text record field. Sentiment analysis is needed to evaluate and understand customer feedback in the record field (Yi and Liu, 2020).

1.1 Research Question

The overarching research question of this study was defined as:

RQ: How could multi-dimensional sentiment analysis on customer feedback be implemented to create an automated response system within Salesforce CRM?

Sub-questions were determined that are connected to the research subject.

SQ1. Which text processing techniques are required to handle unstructured and noisy text content in Salesforce records?
SQ2. Which ML model achieves higher accuracy scores in classification?
SQ3. How can a good connection between the AI API and Salesforce be established?
SQ4. How can the results of the ML model response be displayed in Salesforce?
SQ5. How can actions after getting the sentiment analysis results in Salesforce be handled?

1.2 Research Objective

The research objectives of this study were formulated as:

O1. *Develop a Comprehensive Sentiment Analysis Framework:* Building and implementing a multi-dimensional sentiment analysis framework by leveraging state-of-the-art machine learning algorithms (SVM, RoBERTa, Electra) on customer feedback data.

O2. *Evaluate and Compare Machine Learning Algorithms:* Evaluating and comparing the accuracy of various ML models for multidimensional sentiment analysis on the GoEmotions dataset.

O3. *Enhance Salesforce CRM's Machine Learning Capacity:* Integrating the best-performing sentiment analysis model into Salesforce CRM to enable automatic, real-time sentiment analysis of customer reviews.

O4. *Optimize Text Preprocessing Methods:* Identifying and implementing appropriate text preprocessing techniques that meet the challenges presented by unstructured and noisy text data in Salesforce records.

O5. *Establish Seamless AI-Salesforce Integration:* Providing an effective communication interface between the AI API and Salesforce that ensures seamless data sharing and sentiment assessment operations.

O6. *Design User-Friendly Sentiment Analysis Reporting:* Developing an intuitive and usable interface within Salesforce to display sentiment analysis results to users and enable better decision making.

O7. *Automate Customer Response Actions:* Creating and deploying a system within Salesforce that automatically initiates actions based on sentiment analysis results, thus improving customer response tactics.

2 Related Work

Most sentiment analysis research simply labels text as positive or negative (Mu et al., 2021), without taking into account dimensions of emotion.

2.1 Sentiment Analysis for Better Customer Understanding

Dieksona et al. (2023) investigated if clients are pleased or displeased with Traveloka's services. The dataset was obtained via the Twitter API and includes 1200 tweets regarding Traveloka. The analysis was performed using Python's Scikit-learn module. The study employed three classification methods: SVM, Logistic Regression, and Naive Bayes. The purpose of this study was to evaluate whether the majority of Twitter users consider the functioning of this cell phone travel application as favorable, negative, or neutral. The results showed that SVM is more accurate at identifying the sentiment.

Yi and Liu (2020) used ML models to study, assess, and categorize products and save data based on user experience. Product data and customer input were gathered from a benchmark Unified Computing System (UCS), which is a server for data-driven computing products designed to assess hardware, graphical support, and software management. The findings demonstrated that ML methods outperformed alternative approaches. Compared to other existing systems, the suggested HRS system has a higher MAPE of 96% and an accuracy of around 98%. The proposed HRS system has an average absolute error of around 0.6, indicating that it performs very well.

Qaisar (2020) ran sentiment analysis on IMDB movie reviews using a LSTM. The data was efficiently preprocessed and segregated to enhance classification findings, both positive and negative. The findings show the best categorization accuracy at 89.9%.

Medhat et al. (2014) produced a comprehensive review of sentiment analysis methodologies and applications, highlighting the evolution of sentiment analysis and its use in various fields. They explored various strategies, including ML and their usefulness in evaluating customer sentiment. The study provides a broad comparison of techniques used in sentiment analysis and their effectiveness in various applications.

Joulin et al. (2016) introduced FastText, an effective text classification library widely used for sentiment analysis. The study demonstrated the efficiency and accuracy of FastText in sentiment classification tasks. FastText achieved an accuracy of 92.5% on a large-scale text classification task, with faster training times compared to deep learning models like LSTM and CNN.

Young et al. (2018) presented a comprehensive overview of deep learning methods for sentiment analysis, the study examined various architectures such as CNN, RNN, and attention processes, and their applications in sentiment analysis tasks. While deep learning models such as CNNs and RNNs achieved accuracies ranging from 85% to 92% on various sentiment analysis tasks, attention mechanisms further improved the performance.

Poria et al. (2017) discussed the use of multi-modal sentiment analysis, combining text, audio, and visual data to improve sentiment classification accuracy, which is particularly relevant in customer feedback scenarios where multiple data sources are available. The proposed multi-modal approach achieved an accuracy of 93.4% on sentiment classification tasks involving customer feedback, which is significantly higher than text-only models (89.7%).

2.2 Muti-dimensional Sentiment Analysis

Multidimensional analysis of corpus offers more insight about a corpus and provides better understanding (Gutiérrez-Batista et al., 2021). Comprehending customer emotions from customer feedback gives more insights to the organizations.

Almeida et al. (2011) applied SVM for sentiment analysis on Twitter data. They compared SVM with other classifiers, showing that SVM provided superior performance in terms of accuracy. SVM demonstrated its effectiveness in binary sentiment analysis tasks by achieving an accuracy rate of 97.64% in classifying SMS messages.

Mullen and Collier (2004) used SVM for sentiment analysis in online forums and focused on identifying content as supportive or non-supportive. The study demonstrated the ability of SVM to handle imbalanced datasets. SVM's F1 score was 84.5%, outperforming Naive Bayes and Decision Trees in this domain.

Gan and Yu (2015) conducted sentiment analysis on 268,442 customer reviews from 7,508 restaurants on Yelp.com using five factors: cuisine, service, decor/ambience, special conditions, and pricing. A multilevel model found that opinions on these five parameters alone accounted for around 28% of explainable differences between restaurants and 12% of explainable differences within restaurants in restaurant star ratings. The multi-level model, including extra-level and control variables, explained about 53% of the variance between restaurants and 28% of the variance within them.

Li et al. (2022) used a comprehensive method, incorporating lexicon-based ML sentiment analysis, to gain information on emotion from a corpus. The study found that lexicon-based ML techniques drove other techniques.

Greco and Polli (2020) attempted to create an unsupervised model based on a dataset of brand ratings from customers. A cluster model was created using K-means according to cosine similarity, and a relationship analysis was carried out on the cluster utilizing the keyword matrix to determine sentiment depending on the amount of messages categorized in the group and their meanings. Finally, a network evaluation using the Louvain method was executed. Emotion analysis identified five company representations and attributes for each customer community in terms of item choices (model, color, purchasing possibilities, etc.) and perceptions of the brand (fashion or sports fans).

Bar et al. (2023) emphasized the need of improving customer comprehension. To improve customer comprehension, a model based on the Haar-Cascade Model and CNN was employed to recognize the emotions on different faces and perform multi-dimensional sentiment analysis on client information in real-time. This model accurately predicted 76% of seven emotions. The study's results were used to develop a real-time execution of a more precise and reliable emotion recognition algorithm in business and other applications.

Liu et al. (2019) introduced RoBERTa, a robustly tuned BERT model, and demonstrated its performance on various NLP applications including sentiment analysis. RoBERTa demonstrated state-of-the-art performance on the GLUE benchmark, including the sentiment analysis task with an accuracy of 96.3% on the SST-2 dataset.

Barbieri et al. (2020) investigated the use of RoBERTa for sentiment analysis on Twitter data. The study focused on fine-tuning RoBERTa to respond to different attributes of social media text. The fine-tuned RoBERTa outperformed other models by achieving 90.3% accuracy in sentiment analysis on the TweetEval benchmark.

Clark et al. (2020) introduced Electra as a pre-training strategy to effectively train text encoders and tested its performance on various NLP applications including sentiment analysis. Electra outperformed BERT and RoBERTa on GLUE sentiment analysis tasks, achieving 96.7% accuracy on the SST-2 dataset with a much smaller model size.

Rogers et al. (2021) presented a critical study looking at the strengths and weaknesses of the Electra model in various NLP applications, including sentiment analysis. The study found that Electra outperformed competing models like RoBERTa while using fewer resources and had an accuracy rate of around 96% on traditional sentiment analysis benchmarks.

2.3 CRM Tools and Machine Learning

The majority of the research focuses on out-of-the-box AI solutions for CRMs rather than exploiting ML capabilities. There have been only few studies attempting to harness the ML capabilities of CRM platforms.

Bhosale et al. (2023) endeavored to provide a thorough examination of RPA integration in email automation. The proposed approach aimed to streamline email operations, decrease manual duties, and produce better outcomes. Seamless integration with Salesforce enables the tracking and evaluation of email performance indicators, offering valuable insight into consumer requests. This study utilized UIPath Sentiment Analysis assistance, and the machine learning method achieved a 97% accuracy rate. Automation of AI within Salesforce increased the organization's ability to automate email generation, personalization, organizing, monitoring, and categorization.

Sabbeh (2018) examined and contrasted ML approaches to customer churn prediction in CRM platforms. Discriminant analysis and decision trees (CART), k-nearest neighbors, SVM, logistic regression, ensemble-based learning (random forest, island boosting trees, stochastic gradient boosting), Naive Bayesian, and multi-layer perceptrons were utilized. The algorithms were employed to a telecommunications dataset. Random forest and

ADA Boosting outperformed other algorithms with about the same accuracy score of 96%.

Nguyen and Mutum (2012) investigated the use of data mining techniques in CRM systems to improve customer relationship management. The study demonstrated how ML models may be used to categorize customers, anticipate their behavior, and improve customer retention. The paper presents an overview of several ML applications in CRM, stressing the advantages of data-driven decision-making for customer segmentation and retention strategies.

Nguyen and Simkin (2013) explored the efficacy of ML techniques in CRM systems for customer segmentation and targeted marketing. The study used clustering algorithms to segment clients based on their purchasing behavior and prediction models to optimize marketing campaigns. Ml-based segmentation increased the accuracy of marketing initiatives by 20%, resulting in greater consumer engagement.

Rust and Huang (2014) investigated the impact of ML on customer relationship management in the service business. The study looked at how predictive models can be used to forecast customer lifetime value and improve service delivery. Predictive modeling using ML improved the accuracy of client lifetime value projections by 18%, resulting in more efficient marketing resource allocation.

Ngai et al. (2009) conducted a comprehensive study on the use of data mining techniques in CRM systems, focusing on customer churn prediction in the telecommunications industry. They applied a range of ML methods, including decision trees, neural networks, and SVM, to CRM datasets. The study found that SVM had the highest accuracy in predicting customer churn. The SVM model outperformed the other models in the study, predicting customer churn with 91.2% accuracy. Incorporating these models into CRM systems provided valuable insights into customer retention methods.

3 Methodology

The Cross-Industry Standard Process for Data Mining (CRISP-DM) starts with business goals that are translated into unambiguous data analysis tasks and then executes defined processes, ultimately transforming “data into knowledge” (Martínez-Plumed et al., 2019). The CRISP-DM methodology is shown in Figure 2.

Salesforce offers its own AI tool, Einstein. Einstein is Salesforce’s first sophisticated AI and propels the company to global achievements.[1] Salesforce Einstein can analyze sentiment and classify content as positive, negative, or neutral. Furthermore, Salesforce has limited emotion analysis capacity.[2]

Multi-dimensional sentiment analysis is a more sophisticated method that goes beyond the standard positive, negative, and neutral sentiment classifications. It is effective for applications that demand a more in-depth analysis since it provides a more thorough and nuanced understanding of the sentiment in a given piece of content by focusing on specific features or dimensions. In this project, multi-dimensional sentiment analysis was used on customer feedback to create an effective response mechanism within Salesforce. SVM, RoBERTa, and Electra models were used for sentiment analysis.

The subtasks specify several components and duties for this research, such as integration, automation, NLP, REST API, vectorization, and sentiment analysis. Below, we

[1] https://www.salesforce.com/ap/products/einstein/faq

[2] https://www.salesforcegyan.com/post/sentiment-analysis-using-salesforce

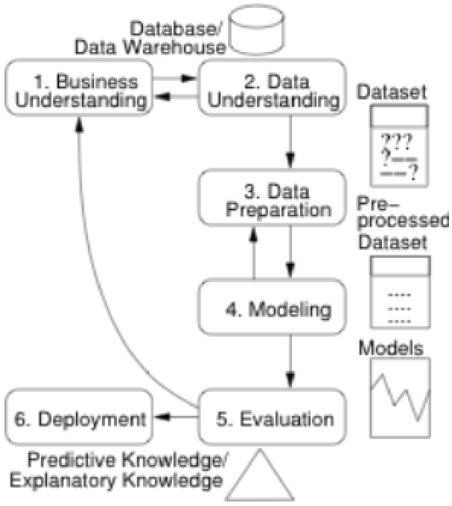

Figure 2: CRISP-DM (Moro et al., 2011).

outline the precise technical procedures and subtasks (ST) that are related to subquestions (SQ) introduced in Section 1 for performing this research.

SQ1. Which text processing techniques are required to handle unstructured and noisy text content in Salesforce records?
ST1. Cleaning noisy text data requires the use of NLP techniques and Python tools such as NLTK.

SQ2. Which ML model achieves higher accuracy scores in classification?
ST2. The accuracy metric of ML algorithms were evaluated to determine which method performed best.

SQ3. How can a good connection between the AI API and Salesforce be established?
ST3. Communication procedures and architectures were reviewed to ensure effective communication.

SQ4. How can the results of the ML model response be displayed in Salesforce?
ST4. A suitable representation technique was selected, such as the generation of an object field or an action.

SQ5. How can actions after getting the sentiment analysis results in Salesforce be handled?
ST5. Based on the sentiment analysis results, a bespoke customer response was given to the consumer, considering the classification.

3.1 Research Resources and Algorithms

The Salesforce developer edition, the Hugging Face platform, open-source ML models, and the open-source GoEmotions dataset were used in this work. The ML models employed in this project are described below.

SVM algorithm: Support Vector Machines (SVM) are supervised machine learning algorithms typically used for classification tasks, but they can also be used to solve regression problems. The main goal of an SVM is to determine the best hyperplane for separating data into classes. In a two-dimensional space, this hyperplane is a line that divides data points of different classes, with the largest gap between the closest points (called support vectors) from each class. SVMs search for the most effective hyperplane for classification. SVs are points that influence the orientation and position of the hyperplane. SVMs work by transforming the input data into a higher-dimensional space using a technique known as kernel trickery, which allows to discover a linear separator even if the data is not linearly separable in its original environment. Popular kernel functions include linear kernels, polynomial kernels, and Radial Basis Function (RBF) kernels (Cortes and Vapnik, 1995). SVMs are most effective when the data is multi-dimensional and the number of data points is limited. It is known for its performance in classification tasks such as text categorization, image recognition, and bioinformatics.

RoBERTa algorithm: The DistilRoBERTa-base model was used in this research. This model is a simplified version of the RoBERTa base model. The model contains 6 layers, 768 dimensions and 12 headings, and has a total of 82 million parameters (compared to 125 million for RoBERTa-base). On average, DistilRoBERTa is twice as fast as RoBERTa-base.[3] DistilRoBERTa can be fine-tuned to perform specific NLP tasks such as sentiment analysis and emotion recognition. Fine-tuning requires stacking task-specific layers on top of the pre-trained DistilRoBERTa model and training it on labeled data for the desired task. DistilRoBERTa creates highly contextualized word embeddings that capture detailed semantic and syntactic information. DistilRoBERTa is capable of performing multi-class sentiment analysis tasks (e.g., negative, positive, and neutral) as well as multi-label sentiment detection tasks. DistilRoBERTa utilizes transfer learning, where information learned during pre-training is applied to subsequent tasks. This helps achieve superior performance.[4]

Electra algorithm: Efficiently Learning an Encoder that Correctly Classifies Token Substitutions (ELECTRA) is a pre-training strategy for language representation developed to increase the efficiency of NLP. Unlike classical models such as BERT, which are based on the masked language modeling (MLM) technique, ELECTRA uses a different mechanism known as substitution token detection. Instead of masking and predicting some tokens, ELECTRA's generator model corrupts the input by replacing some tokens with plausible alternatives. A discriminator model (i.e., the ELECTRA model) is then trained to discriminate between the original and modified tokens. This method is significantly more sample efficient, meaning it uses less computational resources and training time to achieve results comparable to or better than RoBERTa (Clark et al., 2020). ELECTRA is a new approach to self-supervised language representation learning. It can be used to pre-train transformer networks while requiring minimal computation. ELECTRA models are taught to discriminate between "real" and "fake" input tokens generated by another neural network, similar to a GAN discriminator. ELECTRA performs well at small scale, even when trained on a single GPU. At large scale, ELECTRA has been shown to achieve state-of-the-art performance on the SQuAD 2.0 dataset.[5]

[3]https://huggingface.co/distilbert/distilroberta-base
[4]https://huggingface.co/docs/transformers/en/model_doc/roberta
[5]https://huggingface.co/google/electra-base-discriminator

3.2 Research Evaluation

The primary research question is how multi-dimensional sentiment analysis on customer feedback could be implemented to create an automated response system within Salesforce CRM. The accuracy of ML models served as the evaluation criterion for this study's research question. Accurate customer classification leads to an accurate customer response. Furthermore, subquestions were developed as a road map for this research. Each subtask associated with a subquestion represented a milestone in this study.

3.3 Data Selection

This project did not involve human participants. The project made use of the public secondary GoEmotions dataset.[6] This dataset can be used under the permissive Apache License Version 2.0, which allows researchers to use datasets as open source. The ML algorithms which were used in this study were open source and pre-trained. Throughout the project, legal norms and rules, transparency, fairness, and accountability principles were taken into consideration.

3.4 Data Pre-processing

This study used different data pre-processing approaches to prepare the GoEmotions dataset for emotion categorization.

Emoji and special character processing:

1. Emoji Merging: The same emoji sequences were merged into a single sample to avoid repetition.
2. Character Repetition Limitation: Repeated characters were only allowed to be used three times to avoid noise.
3. Special Entity Tokenization: User mentions were replaced with <user> and prices with <price> to facilitate generalization.

Text normalization and cleanup:

1. Lowercase: All text was converted to lowercase to ensure consistency.
2. Stop Word Removal: Common words that contribute little to meaning – except for negations like "not" and "neither" – were eliminated.
3. Lemmatization: Words were reduced to their root form (lemma) to standardize vocabulary.
4. Contraction Expansion: Abbreviations like "can't" were expanded to "cannot" to improve tokenization.

Stop word expansion:

Additional stop words and common terms were added to the stop word list to refine the word cloud visualization.

[6] https://huggingface.co/datasets/google-research-datasets/go_emotions

These pre-processing steps aimed to clean and normalize the text data, making it suitable for feature extraction and model training.

Ekman mapping in emotion classification:

Ekman mapping is a way to categorize a broader range of emotions into a smaller set of basic emotions. It is based on the work of psychologist Paul Ekman, who identified six basic emotions that are universally recognized across cultures: anger, disgust, fear, joy, sadness, and surprise.[7] In this study, Ekman mapping was used to simplify the GoEmotions dataset, which initially contained 28 different emotion labels. For example, emotions like *annoyance* and *irritation* were mapped to *anger*.

There are several reasons why Ekman mapping is useful in emotion classification tasks. It simplifies the problem by reducing the number of classes the model needs to predict (i.e., reduced complexity). By focusing on core emotions, the model may be less susceptible to overfitting to specific nuances of the training data (i.e., improved generalization). Ekman emotions are universally accepted, making the model potentially more applicable to a variety of datasets (i.e., cross-cultural applicability; Oberländer and Klinger (2018)).

In this study, Ekman mapping was applied to indicate the presence or absence of each core emotion for each text sample in the data frame.

3.5 Exploratory Data Analysis (EDA)

In multi-label classification, class imbalance occurs when some labels appear much more frequently in the dataset than others. This can lead to models being biased toward predicting the majority classes and ignoring the minority classes.

Addressing class imbalance is crucial to ensuring that a model learns to correctly predict all labels, regardless of their frequency. Techniques such as oversampling the minority classes, undersampling the majority classes, or using weighted loss functions can help reduce the effects of imbalance and improve overall model performance. Pie charts for the target classes of the final datasets (train, validation, and test) are shown in Figure 3.

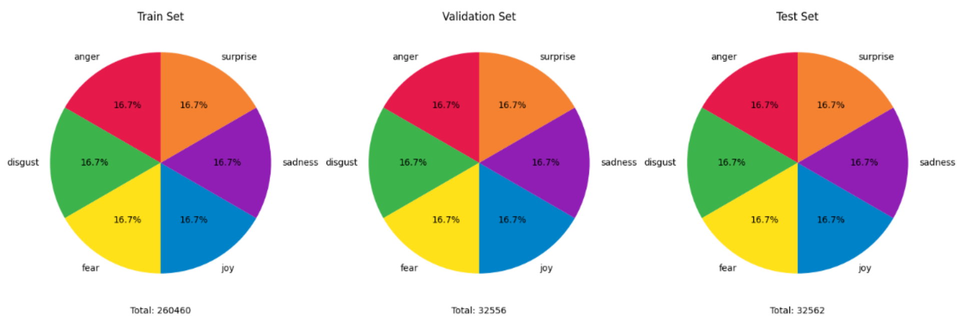

Figure 3: Distribution of target classes (i.e., six basic emotions) in the train, validation, and test data.

Word clouds provide a visual representation of the most frequently used words associated with each tag in a multi-label classification. This can help to quickly identify

[7] https://www.paulekman.com/blog/atlas-of-emotions/

key terms that characterize each class and gain insights into the linguistic patterns that distinguish them.

For example, a word cloud for the tag "joy" might prominently feature words like "happy," "love," or "celebrate," while the tag "anger" might highlight words like "hate," "angry," or "annoyed." This visual analysis can be valuable for understanding the vocabulary that contributes to each tag's prediction and identifying potential biases or misclassifications. Word clouds for the target classes (anger, digust, fear, joy, sadness, and surprise) are shown in Figure 4.

Figure 4: Word clouds of target classes (i.e., six basic emotions) in the data.

4 Design Specification

Different machine learning models (SVM, RoBERTa, Electra) were employed for multidimensional sentiment analysis on the GoEmotions dataset. The best-performing model, Electra, was selected for integration into the final system. The integration involved wrapping the Electra model in a REST API, which allowed it to communicate with Salesforce CRM as an external application. The REST API was developed to serve as an intermediary between the Electra model and Salesforce CRM. It exposed endpoints that Salesforce can call to get sentiment analysis predictions for customer messages. The API was implemented using a Hugging Face Serverless Secure Endpoint[8] framework, and it included routes for sending customer messages to the Electra model and receiving the corresponding emotion label. The API handles pre-processing of input data, model inference, and post-processing of the results before sending them back to Salesforce.

Salesforce CRM was chosen for its robust customer management features and ease of integration with external APIs. The goal was to create an automatic response system

[8] `https://b76gqgdtrr2hmeox.eu-west-1.aws.endpoints.huggingface.cloud`

that could react to customer sentiments in real-time. Salesforce was configured to send customer messages to the REST API and receive the predicted emotion. Based on the received emotion, the CRM automatically triggers predefined responses or actions, such as sending an empathetic message to a frustrated customer or escalating a concern to a human agent. An automatic response system was implemented within Salesforce to handle different emotions. This system included predefined rules and templates for each type of emotion, ensuring that customer interactions were handled appropriately and efficiently.

4.1 Technical Specification

The technical specifications of the present study are given below.

Data Processing: The system required the ability to process and analyze text data efficiently, which is handled by the best performing ML model, i.e., the Electra model.

API Integration: The Hugging Face Serverless Secure Endpoint was used as a REST API to facilitate communication between the ML model and Salesforce CRM.

CRM Integration: Salesforce CRM is capable of interacting with external APIs and triggering automated flows based on the received data. Remote Site Settings is a Salesforce feature that allows users to create a list of external URLs that a Salesforce instance can communicate with. Before making an HTTP call from Salesforce to an external site through Apex code, the endpoint URL should be registered in Remote Site Settings. This is a security feature that prevents unwanted or unsolicited redirects to foreign or potentially dangerous external websites.

Accuracy: The Electra model's predictions maintained a high accuracy level to ensure appropriate responses.

Software: Google Colab Pro was used during the model development, testing, and deployment in Python. The Salesforce CRM edition was the Developer Edition free version with limited memory. APEX, JavaScript, and out-of-the-box features were used for the arrangements in Salesforce. The dataset and the pre-trained models were employed via the Hugging Face platform and its infrastructure provided the Rest API endpoint.

Based on these specifications, the structure/architecture of the project is shown in Figure 5.

5 Implementation

SVM as baseline algorithm, RoBERTa, and Electra models were trained and evaluated using Hugging Face and Google Colab.

5.1 Environment Setup

In this study, Google Colab Pro used as IDE, Hugging Face platform was used for development environment.

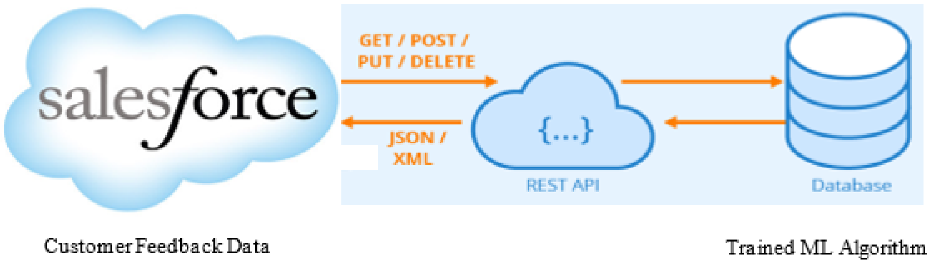

Figure 5: Project architecture.

5.2 Experiment 1 – SVM

An SVM was implemented as a baseline model for sentiment classification for this study. The SVM was trained on the GoEmotions dataset, which contains a broad range of emotion labels. The regularization value C, which handles the trade-off between increasing the margin and lowering the training loss, was set to 0.5. The parameters of the SVM algorithm are shown in Table 1.

Table 1: Parameters for SVM.

Parameter	**Value**
Regularization	0.5
Kernel	Linear, RBF, Polynomial

5.3 Experiment 2 – RoBERTa

RoBERTa is an advanced transformer-based model that builds on BERT by using dynamic masking during training, training on a larger dataset, and for a longer duration. It is particularly well-suited for tasks requiring nuanced understanding of context, such as emotion classification.

The RoBERTa model was fine-tuned on the GoEmotions dataset to perform sentiment analysis. Fine-tuning involved adjusting the pre-trained RoBERTa model's weights on the specific task of emotion classification. Hyper-parameters were optimized during this process to achieve the best possible performance.

The parameters of RoBERTa algorithm are describe in Table 2.

Table 2: Parameters for RoBERTa.

Parameter	**Value**
Evaluation Strategy	Epoch
Training and Evaluation Processes Batch Sizes	16
Number of Epochs	4
Weight Decay	0.01
FR16	True

5.4 Experiment 3 – Electra

The Electra model was fine-tuned on the GoEmotions dataset for sentiment analysis. The fine-tuning step involved adapting the parameters of the pre-trained Electra model to the emotion classification task. During this procedure, hyper-parameters were carefully tuned to improve the performance and accuracy of the model in identifying and classifying sentiments.

The parameters of the Electra algorithm are describe in Table 3.

Table 3: Parameters for Electra.

Parameter	**Value**
Evaluation Strategy	Epoch
Training and Evaluation Processes Batch Sizes	16
Number of Epochs	4
Weight Decay	0.01
FR16	True

5.5 Experiment 4 – AI API and Salesforce Connection

In this study, the *google/electra-base-discriminator* model, which is fine-tuned for emotion classification and hosted in the *barx2os/electra-emotion* repository on Hugging Face, was integrated with Salesforce to perform sentiment analysis on customer feedback.

Model and Endpoint Configuration:

Hugging Face's API was used to install the fine-tuned Electra model, accessible as *electra-emotion-bs*. The endpoint, which is securely accessible over the internet via TLS/SSL, requires a valid Hugging Face token for authentication to ensure that API calls are processed securely (i.e., an authorization token). The endpoint's URL is: `https://b76gqgdtrr2hmeox.eu-west-1.aws.endpoints.huggingface.cloud`.

Before integrating with Salesforce, the endpoint was tested using Postman. The test involved sending HTTP requests with appropriate headers to the endpoint, including the authorization token. The successful responses confirmed the correct functionality of the model in classifying emotions based on the provided input data.

Salesforce Integration:

The Hugging Face API URL was registered under the remote site settings. This step is critical for Salesforce to securely make HTTP calls to external endpoints.

An Apex class was created to manage communication between Salesforce and the Hugging Face API. The class sent HTTP POST queries to the API endpoint with the required data and headers, then processed the model's responses, and finally sent tailored emails to customers.

A screen flow was configured to connect with the Apex class. This flow was used to initiate an API call and present the sentiment analysis results in the Salesforce UI and object fields.

6 Evaluation

The most important aspect of the study is the evaluation metrics that show how effectively the model performs. The three models are evaluated based on their accuracy scores.

6.1 Experiment 1 – SVM

After training, the SVM model was evaluated on a test set and its performance was compared to the other models used in this study. SVM provided a solid baseline, it did not achieve the highest accuracy due to the complexity of the emotion labels in the dataset. While SVM was effective for binary or small multi-class classification tasks, its performance was outpaced by more sophisticated models in large-scale, nuanced tasks like emotion detection. The SVM model accuracy score was 0.82.

6.2 Experiment 2 – RoBERTa

RoBERTa is an advanced transformer-based model that builds on BERT by using dynamic masking during training, training on a larger dataset, and for a longer duration. It is particularly well-suited for tasks requiring nuanced understanding of context, such as emotion classification.

The RoBERTa model was fine-tuned on the GoEmotions dataset to perform multi-label classification. Fine-tuning involved adjusting the pre-trained RoBERTa model's weights on the specific task of emotion classification. Hyper-parameters were optimized during this process to achieve the best possible performance.

The accuracy and loss of the RoBERTa algorithm are shown in in Table 4. The test accuracy score of the RoBERTa model was 0.9153.

Table 4: Accuracy and Loss Scores for RoBERTa.

Epoch	Training Loss	Validation Loss	Accuracy
1	0.251900	0.240167	0.912858
2	0.219100	0.227568	0.916114
3	0.195300	0.222929	0.916022
4	0.181300	0.230773	0.914916

6.3 Experiment 3 – Electra

Electra uses a separate approach called replacement token detection. Instead of masking and guessing specific tokens, Electra's generator model corrupts the input by replacing some tokens with reasonable alternatives.

The Electra model was fine-tuned on the GoEmotions dataset for multi-label classification. Fine-tuning entailed tweaking the pre-trained Electra model's weights for the specific purpose of emotion classification. This procedure involved optimizing hyper-parameters to attain the highest potential performance.

The accuracy and loss of the Electra algorithm are shown in Table 5. The test accuracy score of the Electra model was 0.9198.

Table 5: Accuracy and Loss Scores for Electra.

Epoch	Training Loss	Validation Loss	Accuracy
1	0.217200	0.205120	0.923455
2	0.191900	0.201819	0.924100
3	0.172100	0.202593	0.922871
4	0.162300	0.210130	0.921274

6.4 Experiment 4 – Comparison of Models

The results demonstrated that the Electra model has a higher sentiment categorization accuracy than both SVM and RoBERTa. The comparison of the models' accuracy are shown in Table 6.

Table 6: Comparison of Models.

Model	Accuracy
SVM	0.82
RoBERTa	0.9153
Electra	0.9198

6.5 Experiment 5 – AI API and Salesforce Connection

The connection between the AI API and Salesforce focuses on three key aspects. The accuracy and performance of the sentiment analysis model, the effectiveness of the integration within Salesforce, and the overall user experience within the CRM environment.

Model Performance: The Electra model was tested for its ability to reliably classify sentiments in customer feedback data. Model performance was evaluated using a variety of metrics, including accuracy, precision, recall, and F1 score, where appropriate. The fine-tuned Electra model demonstrated high accuracy in distinguishing sentiments from text input, consistent with early Postman testing results. When integrated with Salesforce, the model maintained its accuracy.

Response Time: API call response time was also tested to verify that the integration offers real-time or near-real-time sentiment analysis. The average response time measured was 1.96 seconds, which is within acceptable limits for CRM operations.

Stability and Reliability: The connection between Salesforce and the Electra model was validated under various scenarios, including varying data volumes and request frequencies. The system handled these conditions flawlessly, and there were no outages during the test period.

The implementation resulted in accurate sentiment analysis and quick response to customers within Salesforce, significantly improving CRM capabilities.

6.6 Discussion

The findings of this study aligned with broader trends observed in sentiment analysis research, particularly the shift toward using more sophisticated deep learning models for fine-grained tasks such as emotion detection. However, the study also highlighted the ongoing challenges of integrating these models into real-world systems such as CRM platforms, where factors such as scalability, latency, and resource constraints must be carefully managed.

The SVM model achieved an accuracy of 0.82, serving as a baseline for more advanced models. While it performed adequately, its limitations in handling complex, multi-class classification tasks were evident, consistent with findings in the literature. Future enhancements could involve exploring non-linear kernels or ensemble methods to improve the performance.

The RoBERTa model outperformed SVM with an accuracy of 0.9153 by leveraging its ability to capture nuanced language details. RoBERTa's accuracy score was close to Electra. RoBERTa's performance was robust and in line with its established success on NLP tasks. This superior performance is attributed to RoBERTa's ability to capture subtle differences in text due to its sophisticated architecture and pre-training on a large corpus.

The Electra model had the best accuracy (0.9198), proving its efficiency and effectiveness in sentiment analysis. The strategy to identify token swaps proved superior, but its performance advantage over RoBERTa was modest. Electra's relatively modest computational requirements made it more suitable given the hardware limitations encountered during the research.

Integrating Electra with Salesforce via a secure API enabled real-time sentiment analysis. However, the integration experienced issues such as occasional delays in API responses, which could impact performance under high traffic conditions. Furthermore, relying on external APIs raises potential security and network reliability concerns. Future research could examine on-premises model deployment within Salesforce to address these concerns.

The implementation resulted in accurate sentiment analysis and quick response to customers within Salesforce, significantly improving the CRM capabilities. However, additional optimizations may be required to address latency and scalability issues to fully leverage the potential of this integration in production.

7 Conclusion and Future Work

This study set out to answer the research question: How could multi-dimensional sentiment analysis on customer feedback be implemented to create an automated response system within Salesforce CRM? The research objectives included testing multiple machine learning models (SVM, RoBERTa, and Electra) for sentiment analysis, integrating the best-performing model with Salesforce, and automating sentiment-based responses to customers.

The results showed that the Electra model outperformed both SVM and RoBERTa with the best accuracy in sentiment categorization. This achievement demonstrated the efficiency and adaptability of Electra for application in resource-constrained contexts. The integration with Salesforce via API was successful and enabled real-time sentiment analysis within a CRM system. However, the study encountered significant hardware and software limitations that increased training times and limited the exploration of more sophisticated designs, especially when training larger models.

Despite these limitations, the study provides important insights into the practical application of AI in enterprise environments, supporting the feasibility of adding advanced NLP models to CRM systems. The study highlighted the importance of carefully considering operational constraints, such as hardware capacity, when deploying complex AI models in production situations.

In contrast to earlier research, this study found that, while cutting-edge models such as Electra and RoBERTa provide greater performance, their practical implementation necessitated careful consideration of the operational environment. This work added to the current body of knowledge by giving numerical evidence on the feasibility and challenges of incorporating advanced NLP models into CRM systems.

Future research should address the hardware limitations found in this study. Increasing computing resources would allow for training and deployment of larger models, such as full-scale RoBERTa or more complex LLMs, potentially increasing the accuracy and efficiency of sentiment analysis.

Furthermore, examining direct integration of LLMs with Salesforce could pave the way for more advanced customer engagement models and potentially automate more difficult customer care tasks. Investigating domain-specific fine-tuning and scalability of these models within Salesforce could provide deeper insights and greater applicability.

In conclusion, this study demonstrated the successful integration of advanced NLP models into Salesforce in order to improve CRM functions.

References

Almeida, T. A., Hidalgo, J. M. G. and Yamakami, A. (2011). Contributions to the study of sms spam filtering: new collection and results, *Proceedings of the 11th ACM symposium on Document engineering*, pp. 259–262.

Bar, A. K., Rout, A. and Chaudhuri, A. K. (2023). Emotica. ai-a customer feedback system using ai, *International Research Journal on Advanced Science Hub* **5**(3): 103–110.

Barbieri, F., Camacho-Collados, J., Neves, L. and Espinosa-Anke, L. (2020). Tweeteval: Unified benchmark and comparative evaluation for tweet classification, *arXiv preprint arXiv:2010.12421* .

Bhosale, S., Dhumal, R., Patkar, V. and Singh, T. (2023). Use of rpa for email automation with salesforce integration, *2023 IEEE 5th International Conference on Cybernetics, Cognition and Machine Learning Applications (ICCCMLA)*, IEEE, pp. 140–144.

Charitha, N. S. L. S., Yasaswi, K., Rakesh, V., Varun, M., Yeswanth, M. and Kiran, J. S. (2023). Comparative study of algorithms for sentiment analysis on imdb movie reviews, *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, Vol. 1, IEEE, pp. 824–828.

Clark, K., Luong, M.-T., Le, Q. V. and Manning, C. D. (2020). Electra: Pre-training text encoders as discriminators rather than generators, *arXiv preprint arXiv:2003.10555* .

Cortes, C. and Vapnik, V. (1995). Support-vector networks, *Machine learning* **20**: 273–297.

Dieksona, Z. A., Prakosoa, M. R. B., Qalby, M. S., Putraa, M., Achmada, S. and Sutoyoa, R. (2023). Sentiment analysis for customer review: Case study of traveloka, *Procedia Computer Science* **216**: 682–690.

Gan, Q. and Yu, Y. (2015). Restaurant rating: Industrial standard and word-of-mouth–a text mining and multi-dimensional sentiment analysis, *2015 48th Hawaii International Conference on System Sciences*, IEEE, pp. 1332–1340.

Gowri, S., Surendran, R., Jabez, J. et al. (2022). Improved sentimental analysis to the movie reviews using naive bayes classifier, *2022 International Conference on Electronics and Renewable Systems (ICEARS)*, IEEE, pp. 1831–1836.

Greco, F. and Polli, A. (2020). Emotional text mining: Customer profiling in brand management, *International Journal of Information Management* **51**: 101934.

Gutiérrez-Batista, K., Vila, M.-A. and Martin-Bautista, M. J. (2021). Building a fuzzy sentiment dimension for multidimensional analysis in social networks, *Applied Soft Computing* **108**: 107390.

Joulin, A., Grave, E., Bojanowski, P. and Mikolov, T. (2016). Bag of tricks for efficient text classification, *arXiv preprint arXiv:1607.01759* .

Li, H., Chen, Q., Zhong, Z., Gong, R. and Han, G. (2022). E-word of mouth sentiment analysis for user behavior studies, *Information Processing & Management* **59**(1): 102784.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. and Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach, *arXiv preprint arXiv:1907.11692* .

Martínez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernández-Orallo, J., Kull, M., Lachiche, N., Ramírez-Quintana, M. J. and Flach, P. (2019). Crisp-dm twenty years later: From data mining processes to data science trajectories, *IEEE transactions on knowledge and data engineering* **33**(8): 3048–3061.

Medhat, W., Hassan, A. and Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey, *Ain Shams engineering journal* **5**(4): 1093–1113.

Moro, S., Laureano, R., Cortez, P. et al. (2011). Using data mining for bank direct marketing: An application of the crisp-dm methodology, *Proceedings of the European Simulation and Modelling Conference-ESM*, Vol. 2011.

Mu, R., Zheng, Y., Zhang, K. and Zhang, Y. (2021). Research on customer satisfaction based on multidimensional analysis, *International Journal of Computational Intelligence Systems* **14**(1): 605–616.

Mullen, T. and Collier, N. (2004). Sentiment analysis using support vector machines with diverse information sources, *Proceedings of the 2004 conference on empirical methods in natural language processing*, pp. 412–418.

Ngai, E. W., Xiu, L. and Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification, *Expert systems with applications* **36**(2): 2592–2602.

Nguyen, B. and Mutum, D. S. (2012). A review of customer relationship management: successes, advances, pitfalls and futures, *Business Process Management Journal* **18**(3): 400–419.

Nguyen, B. and Simkin, L. (2013). The dark side of crm: advantaged and disadvantaged customers, *Journal of Consumer Marketing* **30**(1): 17–30.

Oberländer, L. A. M. and Klinger, R. (2018). An analysis of annotated corpora for emotion classification in text, *Proceedings of the 27th international conference on computational linguistics*, pp. 2104–2119.

Poria, S., Cambria, E., Hazarika, D., Mazumder, N., Zadeh, A. and Morency, L.-P. (2017). Multi-level multiple attentions for contextual multimodal sentiment analysis, *2017 IEEE International Conference on Data Mining (ICDM)*, IEEE, pp. 1033–1038.

Qaisar, S. M. (2020). Sentiment analysis of imdb movie reviews using long short-term memory, *2020 2nd International Conference on Computer and Information Sciences (ICCIS)*, IEEE, pp. 1–4.

Rogers, A., Kovaleva, O. and Rumshisky, A. (2021). A primer in bertology: What we know about how bert works, *Transactions of the Association for Computational Linguistics* **8**: 842–866.

Rust, R. T. and Huang, M.-H. (2014). The service revolution and the transformation of marketing science, *Marketing Science* **33**(2): 206–221.

Sabbeh, S. F. (2018). Machine-learning techniques for customer retention: A comparative study, *International Journal of advanced computer Science and applications* **9**(2).

Yi, S. and Liu, X. (2020). Machine learning based customer sentiment analysis for recommending shoppers, shops based on customers' review, *Complex & Intelligent Systems* **6**(3): 621–634.

Young, T., Hazarika, D., Poria, S. and Cambria, E. (2018). Recent trends in deep learning based natural language processing, *ieee Computational intelligenCe magazine* **13**(3): 55–75.