Generate sales leads from email conversation using AI/ML

DISSERTATION

Submitted in partial fulfillment of the requirements of
M. Tech. Software Engineering Degree programme

By

Roshan Kumar GUPTA
ID No. 2018HS70003

Under the supervision of:

Rupali ARORA
&
M Saurabh SHARMA

Dissertation work carried out at
SAP Labs, Bangalore

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI, PILANI CAMPUS
April 2022

SE SAP ZG629T DISSERTATION

Generate sales leads from email conversation using AI/ML

Submitted in partial fulfillment of the requirements of
M. Tech. Software Engineering Degree programme

By

Roshan Kumar Gupta
ID No. 2018HS70003

Under the supervision of:

Rupali Arora
&
M Saurabh Sharma

Dissertation work carried out at
SAP Labs, Bangalore

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI, PILANI CAMPUS
April 2022

Acknowledgements

I would like to acknowledge and extend my heartfelt gratitude to Prof. Anitha N, My SAP Labs project supervisor had volunteered to provide essential help and advice in the development of this suggested project. Mrs. Rupali Arora and Mr. Saurabh Sharma deserve special mention for their help and guidance with project feedback and examinations. Without their encouragement and assistance, I would not have been able to finish my job.

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
SECOND SEMESTER 2021-22

SESAP ZG629T DISSERTATION

Dissertation Title: Generate sales leads from email conversation using AI/ML

Name of the Supervisor: Rupali Arora

Name of the Student: Roshan Kumar Gupta

BITS Id of the Student: 2018HS70003

Abstract

The goal of this project is to develop a highly intelligent machine learning model that can predict sentiment from sales emails. The goal is to provide a working UI dashboard with email template creation and a machine learning model. Users can input data into the UI, and the machine learning model will analyse the data and make predictions that will aid salespeople in making decisions. It automates the procedure and decreases the time it takes for manual processes.

Keywords: Machine Learning, Supervised Learning, Unsupervised Learning, Natural language processing.

Abbreviations

UI	**U**ser **I**nterface
URI	**U**niform **R**esource **I**dentifier
API	**A**pplication **P**rogramming **I**nterface
ML	**M**achine **L**earning
NLP	**N**atural **L**anguage **P**rocessing

List of Tables

Table 9.1 – Complete Project Plan v
Table 10.1– Work Accomplished v

List of Figures

7.1 Architecture of Machine Learning Model 8
7.2 Sequence diagram of ML solution 10

Contents

Acknowledgements i

Abstract ii

Abbreviations iii

List of Tables iv

List of Figures v

Contents vi

1 Introduction & Background 1

2 Problem Statement 3

3 Objective of the Project 4

4 Uniqueness of the Project 5

5 Benefit to the Organization 6

6 Scope of Work 7

7 Solution Architecture 8

8 Resources Needed 11

9 Project Plan and Deliverable 12

10 Work Accomplished so far 14

11 Key Challenges Faced 16

12 Potential risks and mitigation plan 17

13 Plan for remainder of the project 18

Bibliography 19

Chapter 1

Introduction & Background

The project is centered around the Natural language processing. NLP is a field of Artificial Intelligence that gives the machines the ability to read, understand and derive meaning from human languages. It can solve various problems related to natural languages like sentiment analysis, speech recognition and etc.

The process of detecting positive or negative sentiment in text is known as sentiment analysis. Businesses frequently utilise it to detect sentiment in social data, assess brand reputation, and gain a better understanding of their customers. Sentiment analysis is quickly becoming a crucial tool for monitoring and understanding sentiment in all forms of data, as humans communicate their thoughts and feelings more openly than ever before. Brands can learn what makes customers happy or frustrated by automatically evaluating customer feedback, such as comments in survey replies and social media dialogues. This allows them to customise products and services to match their customers' demands.

The sales process, often known as the sales cycle, is the process by which your organisation sells its product or service to clients. From the initial contact with a lead through the final sale, it entails a sequence of processes. There are seven stages in the sales process but this project mainly deals with identification of leads. Prospect for potential customers.

- Make contact with prospects.
- Qualify prospects.
- Nurture prospects.

- Present your offer.
- Overcome objections.
- Close the deal.

Chapter 2

Problem Statement

A sales lead is a person or business who may eventually become a client. Sales lead also refers to the data that identifies an entity as a potential buyer of a product or service. Businesses gain access to sales leads through advertising, trade shows, direct mailings, third parties, and other marketing efforts. A sales lead is not really a sales "prospect" per se because a business would need to examine and qualify the potential new client further to determine their intent and interest.

There are various types of sales leads, such as hot leads, cold leads, warm leads, information qualified leads, and so on. Salespeople spent more time reading marketing and campaign emails in order to discover leads who could become potential customers. The primary goal of this project is to eliminate the amount of time salespeople spend identifying sales leads.

Chapter 3

Objective of the Project

The main objective and outcome this project is to train sales data with both supervised and unsupervised alogorithm, it will also extract named entity information from mail contents which can give vital information to end users and develop a rest API using python and flask for the inference. There will be an machine learning solution that will assist salespeople in identifying leads more quickly and saving time.

There will be UI dashboard made of UI5 framework where users can see the results and visualizations in graph.

The UI will also allow generating the template for sending emails to customer, Analysis on the possibility of integration with outlook. it solves the problem by reducing the the time and effort of sales people and allow sales team to sell products more effectively.

Chapter 4

Uniqueness of the Project

SAP already has a product for handling sales leads and sales processes, and it includes a number of distinct processes. However, it is primarily manual. This type of activity is also done manually by sales teams. This initiative eliminates human labour while simultaneously making the process more intelligent through the use of machine learning. It is a narrow sector that only involves email communication. The dataset has been trained using both supervised and unsupervised algorithms, making the ML model more clever when it comes to identifying the sentiment from email conversations.

Chapter 5

Benefit to the Organization

SAP has a solution for this called lead management, which controls the entire sales process, but it is a manual procedure. In SAP, there is a sales staff that keeps track of leads in excel and manually identifies sales leads. These are the benefits it provides to salespeople in SAP.

- Reduce the time and effort for salesperson.
- ML model to assist sales team.
- Genereate named entity recognition from emails
- UI dashboard to display the model results and leads information.
- Generate a template for user which can be used for sending a mail.

Chapter 6

Scope of Work

The scope of this project is to develop a machine learning model using supervised(SVN, random forest, naive bayes) and unsupervised(text blob) earning classification algorithm and train it on a sales dataset that can be used to predict various types of sales leads. The ML nodel will also construct named entity recognition, which can be useful for decision making. Customers will be able to study the results and graph visualisations using a UI dashboard developed on the UI5 framework. The user interface will also let you create a template for delivering emails to your consumers.

Chapter 7

Solution Architecture

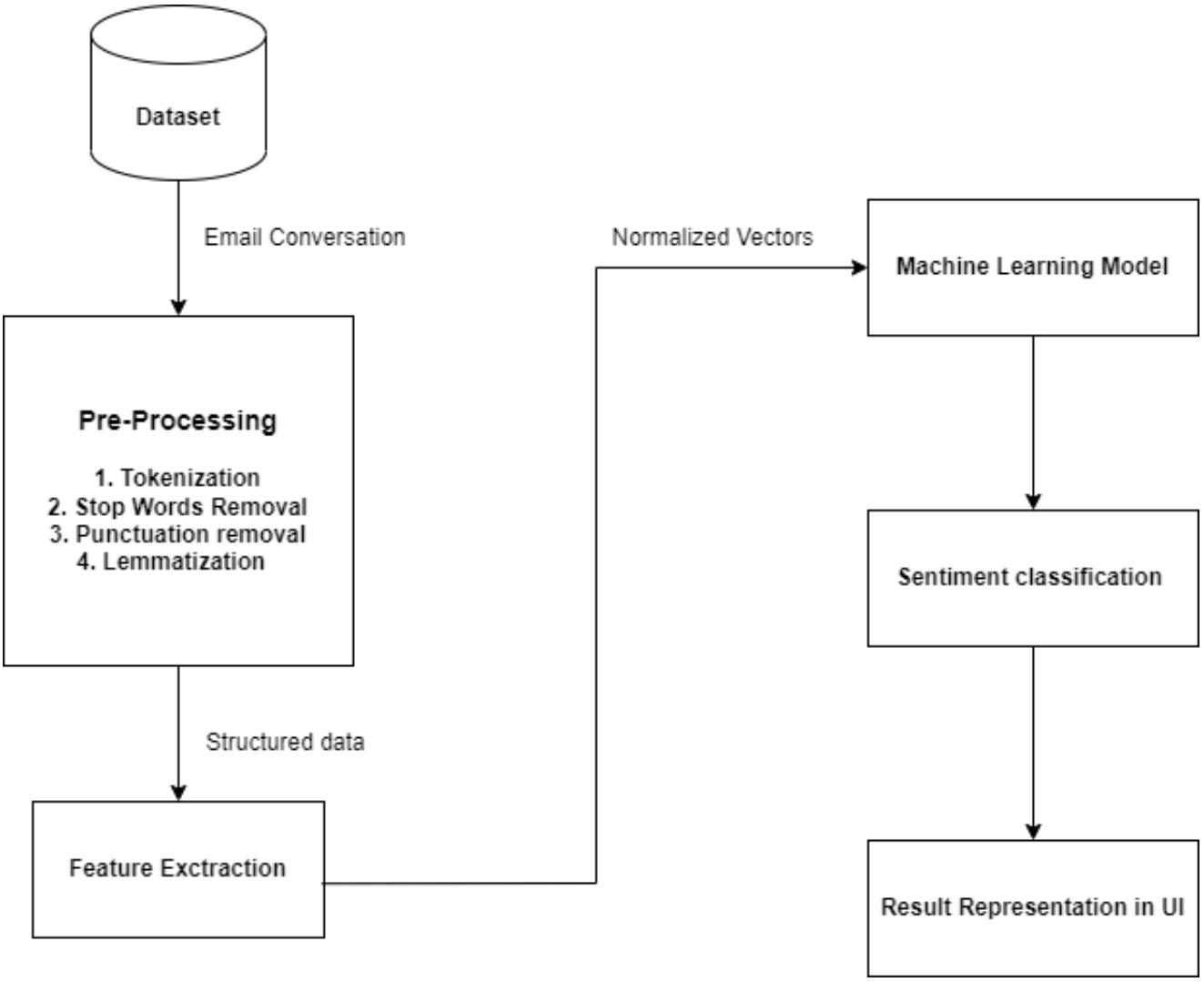

FIGURE 7.1: Architecture of machine learning model

The proposed approach can be broken down into many phases.

Dataset

The first step is to get data from various sources or to create it yourself. For this project, the majority of the dataset was obtained through Kaggle, and the data needed to be labelled in order to train with supervised algorithms.

Pre-Processing

The data is usually in text format, so it will be pre-processed and cleaned in this step, such as removing punctuation, stopwords, and special characters, stemming, leammatization, and tokenization, and the output will be clean text. To pre-process the text, NLTK and Spacy python libraries were used.

Feature Extraction

The important features of the document will be extracted in this stage, and the token will be turned into a numerical value. Because only numeric values are recognised by the machine. Count vecorization and TF-IDF are two methods that have been utilised to turn tokens into vectors.

Machine Learning Model

After the data has been cleaned and normalised into vectors, the next step is to train the data using machine learning algorithms. ML algorithms that have been utilised include Naive Bayes, Random Forest, and SVM, among others. Lexicon-based techniques have also been employed for sentiment analysis.

Sentiment Classification

In this step, a trained ML model was serialised into pickel format using SKlearn, and a rest API was built on top of the serialised model for prediction and UI integration.

UI Representation

The model predictions were utilised to construct a user interface dashboard that allows users to view analytics and visualisations.

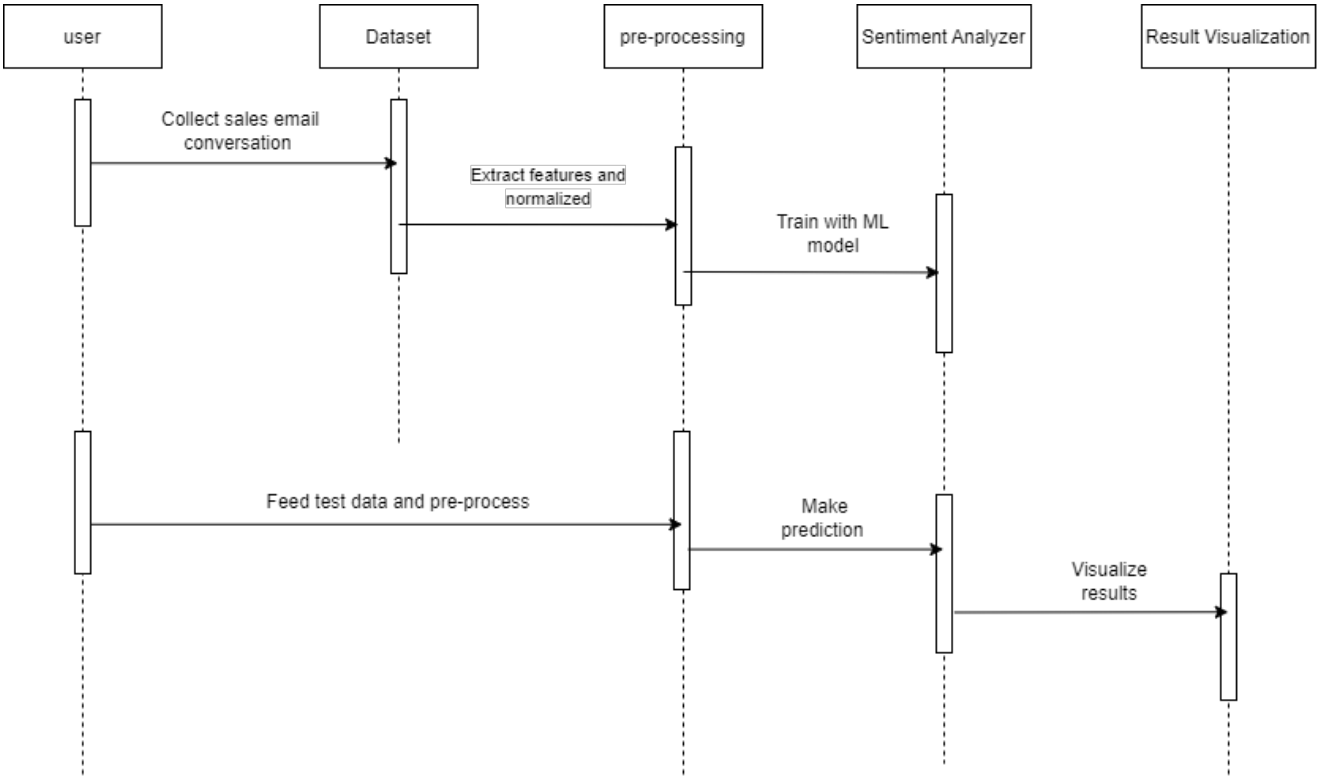

FIGURE 7.2: Sequence diagram of ML solution

Chapter 8

Resources Needed

- Python Server
- Sales data-set
- Jupyter Notebook
- UI5
- Outlook

Chapter 9

Project Plan and Deliverable

Serial. Number	Tasks	Start date - End date (in Calendar Weeks)	Planned duration (weeks)	Specific deliverables
1.	System and database set up for creating machine learning model and API	21 Feb to 27 Feb	1	Complete system and database set up for development of project
2.	Analyze and collect sales conversation data for training the machine learning model	28 Feb to 13 Mar	2	Complete sales data set for training the algorithm
3.	Analyze different ML algorithms(supervised and unsupervised), Implement ML model and train the model with data set	14 Mar to 17 Apr	5	Trained machine learning model with data set and named entity recognition

4.	Hyper parameter optimization of algorithm	18 Apr to 24 Apr	1	Increased Accuracy of algorithm
5.	Create a rest API for the model which can be consumed	25 Apr to 1 May	1	Rest API which can be used for UI integration
6.	Create UI dashboard for displaying model results and relevant information about leads	2 May to 22 May	3	UI dashboard for displaying model results and visualizations
7.	Generate a template which can used by customers for response	23 May to 5 Jun	2	Developing a email template
8.	Analysis on possibility of integration with outlook	6 Jun to 12 Jun	1	Analysis result

Chapter 10

Work Accomplished so far

Serial. Number	**Tasks**	**Start date - End date (in Calendar Weeks)**	**Planned duration (weeks)**	**Specific deliverables**
1.	System and database set up for creating machine learning model and API	21 Feb to 27 Feb	1	Complete system and database set up for development of project
2.	Analyze and collect sales conversation data for training the machine learning model	28 Feb to 13 Mar	2	Complete sales data set for training the algorithm
3.	Analyze different ML algorithms(supervised and unsupervised), Implement ML model and train the model with data set	14 Mar to 17 Apr	5	Trained machine learning model with data set and named entity recognition

4.	Hyper parameter optimization of algorithm	18 Apr to 24 Apr	1	Increased Accuracy of algorithm
5.	Create a rest API for the model which can be consumed	25 Apr to 1 May	1	Rest API which can be used for UI integration and basic UI App

Chapter 11

Key Challenges Faced

The key challenges of this project so far have been:

- To get the correct labeled dataset for sales email conversation
- To get good accuracy from suppervised learning algorithm
- To design a UI for email template generation
- Integrating the app with outlook
- Designing the UI for data visualisations

Chapter 12

Potential risks and mitigation plan

The most major risk identified thus far has been linked to supervised model training accuracy, which indicates that the accuracy acquired will be poor if data is not labelled correctly or there is insufficient data for training. The training data must be accurately labelled and in sufficient volumes to achieve high accuracy.

Chapter 13

Plan for remainder of the project

Serial. Number	**Tasks**	**Start date - End date (in Calendar Weeks)**	**Planned duration (weeks)**	**Specific deliverables**
6.	Create UI dashboard for displaying model results and relevant information about leads	2 May to 22 May	3	UI dashboard for displaying model results and visualizations
7.	Generate a template which can used by customers for response	23 May to 5 Jun	2	Developing a email template
8.	Analysis on possibility of integration with outlook	6 Jun to 12 Jun	1	Analysis result

Bibliography

[1] Natural Language Processing

https://link.springer.com/chapter/10.1007/978-81-322-3972-7_19

[2] Sales Leads

https://www.investopedia.com/terms/s/sales-lead.asp

[3] Rule Based Sentiment Analysis

https://www.analyticsvidhya.com/blog/2021/06/rule-based-sentiment-analysis-in-python