REVA UNIVERSITY
Bengaluru, India

A Project Report on

# Automating Customer Experience Audit using Per-trained Generalized Models

**Submitted in Partial Fulfilment for Award of Degree of**
**Master of Business Administration**
**In Business Analytics**


**Submitted By**
**Saumyadip Sarkar**
R19MBA07

**Under the Guidance of**
**Ravi Shukla**
Consultant, Data Science at Dell Technologies

REVA Academy for Corporate Excellence - RACE
REVA University
Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bengaluru - 560 064
race.reva.edu.in


**February, 2021**

REVA UNIVERSITY
Bengaluru, India

# Candidate's Declaration

I, Saumyadip Sarkar hereby declare that I have completed the project work towards the Master of Business Administration in Business Analytics at REVA University on the topic entitled 'Automating Customer Experience Audit using Per-trained Generalized Models' under the supervision of Ravi Shukla Consultant, Data Science at Dell Technologies. This report embodies the original work done by me in partial fulfilment of the requirements for the award of degree for the academic year 2020.

Saumyadip Sarkar

Place: Bengaluru — Saumyadip Sarkar

Date: 27-Feb-2021 — Signature of Student

# Certificate

This is to Certify that the project work entitled ‘Automating Customer Experience Audit using Per-trained Generalized Models’ carried out by Saumyadip Sarkar with SRN R19MBA07, is a bonafide student of REVA University, is submitting the project report in fulfilment for the award of MBA in Business Analytics during the academic year 2020. The Project report has been tested for plagiarism and has passed the plagiarism test with the similarity score less than 15%. The project report has been approved as it satisfies the academic requirements in respect of PROJECT work prescribed for the said Degree.

| | |
|---|---|
| Ravi Shukla | Dr. Shinu Abhi |
| Guide | Director |

External Viva

Names of the Examiners

1. Indrajit Kar, Head of AI-Chief Architect & Data Scientists, Siemens
2. Pradeepta Mishra, Associate Principal & Head of AI, LTI -Larsen and Toubro Infotech.

Place: Bengaluru

Date: 02-Mar-2021

REVA
UNIVERSITY
Bengaluru, India

# Acknowledgement


I would like to express my sincere gratitude to our Chancellor **Dr. P. Shyama Raju**, **Dr. S.Y. Kulkarni**, Ex-Vice Chancellor, **Dr. D.K. Mallikharjuna Babu**, Vice-Chancellor **and Dr. Dhananjaya**, Pro Vice-Chancellor for supporting the RACE program, specifically designed for working professionals, and providing necessary facilities and infrastructure required for the best learning experience. I am immensely proud of being part of this program and REVA University.

I would like to thank **Dr. Shinu Abhi**, Director, RACE (REVA Academics for Corporate Excellence) for her guidance and constant supervision in providing necessary information regarding the project thus ensuring complete adherence to project guidelines.

A special thanks to my project guide **Mr. Ravi Shukla** for his valuable guidance provided in understanding the concept and executing the project.

I am also thankful to my fellow classmates for their constructive criticism and friendly advice during the project work.

Last but not the least, I would like to thank my family and colleagues who have extended their invaluable co-operation and encouragement throughout this project lifecycle.

Place: Bengaluru

Date: 27-Feb-2021

Name of the Student:

Saumyadip Sarkar

REVA UNIVERSITY
Bengaluru, India

# Similarity Index Report

This is to certify that this project report titled **Automating Customer Experience Audit using Per-trained Generalized Models** was scanned for similarity detection. Process and outcome is given below.

Software Used: Turnintn
Date of Report Generation: 02-Mar-2021
Similarity Index in %: 7%
Total word count: 5938
Name of the Guide: Ravi Shukla

Saumyadip Sarkar

Place: Bengaluru
Date: 02-Mar-2021

Saumyadip Sarkar
Signature of Student

Verified by: Andrea Brian C

Signature
Dr. Shinu Abhi,
Director, Corporate Training

## List of Abbreviations

| Sl. No | Abbreviation | Long Form |
|---|---|---|
| 1 | NLU | Natural Language Understanding |
| 2 | BERT | Bidirectional Encoder Representations from Transformers |
| 3 | NLP | Natural Language Processing |
| 4 | NLG | Natural Language Generation |
| 5 | CRSIP-DM | Cross Industry Process for Data Mining |
| 5 | RoBERTa | A Robustly Optimized BERT Pretraining Approach |
| 6 | POS | Part of Speech |

## List of Figures



## List of Tables

# Abstract

A specific organization has been transforming itself from being product centric to customer centric. Taking this approach forward, its finance department has been relentlessly working on resolving customer's dispute related to invoice generation. Customers would approach with their concerns on Invoices through an email interface and all further communications happen through this mail platform. The quality audit team is tasked to identify whether the response from the organization fulfils certain business parameters. In order to fulfil these checks, the quality audit team is currently working on sampling approach and a team of auditors manually read through each identified sample to check whether each response follows the business defined parameters. This is a time-consuming affair and post audit, a humongous amount of time is also spent on calibration between the Auditors and master calibrator. There always remains a grey area when performing such task manually.

The aim of this study is to propose an automated audit check solution using a mix of approaches like Text Mining, Text Cleaning, Natural Language Processing, POS Identification, Pattern Identification, Question-Answer pipeline, and Sentiment Analysis using pre-trained models available in Open-Source platforms. This approach is adopted due to unavailability of huge amount required data to build and train a model on unstructured text.

This would greatly reduce the manual effort of reading and auditing and can be applied to entire population rather than only a few samples.

This proposed approach of auditing quality of response may not be limited to a particular organization but can also be extended to all such organizations who want to check how they are responding to customer queries.



***Keywords: Unstructured Data, Text Mining, Text Cleaning, Natural Language Processing, POS Identification, Pattern Identification, Question-Answer pipeline, Sentiment Analysis, Open Source, BERT***

# Contents

# Chapter 1: Introduction

Of late, organizations have realised the importance of customer satisfaction. It's the customer, who can make or break an organizations. So, delighting cutomers with world class customer service has become paramount importance even for product centric organizations, It's also important to have a culture of customer first approach throughout an organizations. Having said that, the important thing to consider here is how can we measure customer satisfaction. One and the most popular way of doing this is by customer survey. With recent cutting edge development in the space of text analytics and natural langugae processing, we are able to understand the sentiment of customer response very efficiently. However, it is also crtically important to measure how organizations are resolving customer queries. Is there a way to measure this?

The hypothesis here is, if we are able to reslove customer queries efficiently following certain pre-defined rules / criterias, we will be able to delight our customers which in turn will lead to greater customer satisfaction.

Taking this concept forward, the finance department, who is reponsible for resolving queries related to customer invoices, has come out with a audit based approach. Samples are collected from a population of customer quries where there is atleast on response from this department, which then are audited against a set of pre-defined parameters and scored based on hits and misses. This involves reading through each mails by auditors and marking these against audit parameters. Though most of the audit parameters are very specific to this department, for the scope of this study, we have only those audits are considered which are applicable to most other customer centric departments as well.

In recent times, we have seen a lot of advancement in the field of natural language processing, notably Transfomers based state of the art NLP processing for Pytorch and TensorFlow 2.0. It provides pretrained models built using architectures like BERT, GPT-2, RoBERTa, etc.for natural language processing and natural languange generation. With more than 1000 of pretained models available, a generalize approach of solving business problem specific to a particular organization can be adopted. Yes, the accuracy rate may not be as close to building our own models, however in the present context, we have relied on a few of these pre-trained models and used a combination of these to solve the problem at hand.

This study aims to aid auditors with a solution that would help them take better and mostly error free decisions. This will also help them increase their scope of audits and provide a holistic picture of responses being sent against customer queries. This would also bring in efficiency in terms of man hour savings.

# Chapter 2: Literature Review

Natural Langunage processing, undertanding and generation have been an area of interest for decades but with recent development of state-of-the art deep learning models, we are close to achieving human level efficiecny even when the nature of the data is very unstructured.

One such recent development is ‘the Transformers’, a general puprose architecture (BERT, GPT-2, RoBERTa, XLM, DistilBert, XLNet…) for NLP and NLU and NLG. It contains more than 32 pre-trained models and has deep interoperability with TensorFlow 2.0 and PyTorch.(Wolf et al., 2019). It allows us a quick way of leveraging pipeline API using only a few lines of codes. The quickest way of using a pre-trained model on a given data is to use out-of-the box pipeline.(*Quick Tour — Transformers 3.1.0 Documentation*, n.d.).

We have also seen studies on question answering pipeline using unsupervised approach. Once such study was done for solving fact-checking problem which was able to verify facts with right kind of factoid questions (Jobanputra, 2019).

However, the most widely used of transformer models is ‘Bidirectional Encoder Representations from Transformers’ commonly know as BERT. A study on how BERT works reveals some interesting features like interpretable information is actually found within the hidden states of Transformer models which can further be used in identifying weakness of the model like misclassification (van Aken et al., 2019).

Another version of BERT that is gaining immense popularity is DistilBERT. Is is lighter, smaller and faster version of BERT. Study done by Huggingpost shows that it is possible to decrease the size of a BERT model by almost 40%, while maintaining 97% of its language understanding capabilities. At the same time it is almost 60% faster (Sanh et al., 2019). This makes DistilBERT suitable for rapid use in many areas – one of which we have tried to explore in this study.

As we drill more into the transformer models we see a very simple pipeline doing all the heavy lifting task like Tokenizer definition →Tokenization of Documents →Model Definition

→Model Training →Inference when dealing with unstructure data like text. (*Working with Hugging Face Transformers and TF 2.0 | by Akash Desarda | Towards Data Science*, n.d.).

A great use of DistilBERT is nicely put across by Jay Alammar in his blog (*A Visual Guide to Using BERT for the First Time – Jay Alammar – Visualizing Machine Learning One Concept at a Time.*, n.d.). He makes use of DistilBERT heavy lifting capability to generate sentence embedding and then take the result of DistilBERT's processing into a logistic regression model to clasify as positive or negative.

The main advantage of using BERT architecture is that it is possible to deal with relationships between distant words better than recurring networks thus making it versatile and very powerful.(M'Haimdat, n.d.).

As seen earlier, BERT uses a part of transformer architecture which is the first transduction model based entirely on attention. It successfully replaces the recurrent layers commonly used encoder-decoder architectures with multi-headed self-attention. (Vaswani et al., 2017). This makes training of transformer based model faster and paved way for more recent development of pre-trained models based on transformer architecture.

However one study seem to have a different view on attention mechanism. It highlights the fact that relations between weights and model output is not very clear (Wiegreffe & Pinter, 2020). This paved the way another interesting study done to analysze BERT models and inparticular regarding the question answer technique. It also provides the option to discover the internal state of a model at each layer. This approach is termed as VisBERT(Aken et al., 2020).
Still, if we consider BERT and it's phenomenal success in recent times, it's due its versatile nature providing enough scope of fine-tuning opprtunities to enhnace the performance in text classification task (Sun et al., 2019).

Another very important development in the area of NLP is rule-based matching. spaCy's rule-based matcher engine is worth exploring here. It lets us quickly find words and phrases we are looking for and also gives us access to the tokens within the document with their relationships. (*Rule-Based Matching · SpaCy Usage Documentation*, n.d.)

# Chapter 3: Problem Statement

Currently, an auditor is assigned a set of samples picked from a population of responses sent by internal team who is tasked in resolving customer queries. The auditors would open each such mails, read through the queries (it can go up to 10 to 15 to-and-fro responses at times), perform audits based on parameters which are divided into sub-parameters and corresponding checkpoints. Each parameter contains a specific question that need to be checked while performing the audits.

This entire activity of auditing is done manually by reading through each response, interpreting against pre-defined parameters, and marking them as hit, misses or observations. Each audit approximately takes 10 to 20 minutes depending on the complexity of the query thereby limiting the ability to go beyond 20 audits per day per auditors considering 6.5 hours of effective working hours.

Post audit, there is a calibration session done between master auditors and the auditors which happen to be very subjective in nature and thus giving rise to a lot of ambiguity on the findings. Analyst who are marked down often contest the findings of the auditors. Thus, this entire process is very manually intensive and exhausting. Also, the approach is sample based thus limiting our ability to have an overall picture of the quality of response.

The other problem is when an auditor moves out of the process. In such case, we have no other choice but to reduce our sample size. Also, a new auditor is usually given 3 to 4 months of intensive training before being put into the audit process thus creating a huge void during this onboarding phase. Overall, this process at present is very much human dependent.

# Chapter 4: Objectives of the Study

Considering the present problem area, this study aims to identify a set of parameters which can be tackled through natural language processing technique. We would like to explore available pre-trained models, identify the ones that would solve the problem to arrive at the desired result.

We also expect this study would pave way for further development in this field by creating a data repository on which a model can be trained. The whole objective of this study is to minimize the manual intervention by auditors while doing audits and present them with only those audits where our present approach may not have given desired outcome.

Also, through this approach, we aim to solve a unique business problem of checking the quality of response sent against the customer queries. This is usually not the case when we think of customer service. We, most of the time, go for identifying the sentiments of customers through their responses without realizing the fact that the customer sentiment would most likely be driven by how we approach customer requests or queries.

The idea here is fix the root by checking if we can connect with our customers and have an empathetic approach while trying to resolve their queries. All this while, we are doing this through human intervention and interpretation by manually checking the quality of responses.

However, with this study we aim to partially replace this manual intervention using a combination of pre-trained NLP models. We may be not able to tackle all the audit parameters with this approach since we are using general purpose architecture, however it is a new way of looking at a business problem.  This would not only ensure a top-notch customer centric approach but would help build a culture of delighting customers.

# Chapter 5: Project Methodology

This project uses CRSIP-DM framework which begins with understanding the business as a whole and then narrowing it down the specific area.

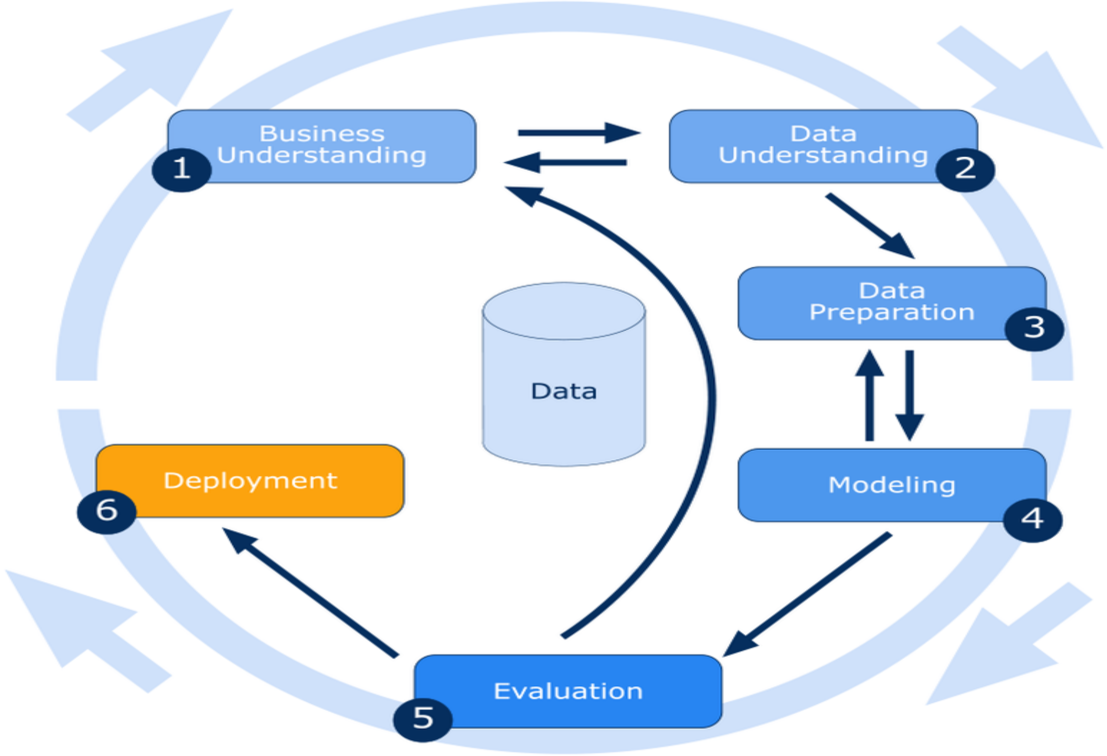


Fig.5.1 – **CRISP-DM Framework**

In this case, it is about quality audit team responsible for identifying the quality of responses sent while resolving a customer query. The quality team is given a set of parameters against which each sampled response is marked.

We then move to data understanding. Here, the data is mostly in form of text generated in from of mail communications. The extraction process is manually done by downloading all the sampled responses from a mail client.

Once the data is collected, we move to data preparation which involves identifying customer queries and corresponding responses. We have used a python script to clean and extract the data. So far, we can retrieve around 80% of data in desired format.

We then follow a mix of approach in the data modelling phase. We have used question-answer pipeline available in transformers model to score responses against the context (queries). This approach gives us desirable result when the response is very straight through against a query, however when the response is not very direct in nature, it may show as “key error” meaning against the context, the model could not identify any direct answer. We have used this logic to identify how a response is against the context. The other model we have used is sentiment analysis pipeline, again available in transformers to understand the use of tonality in the responses.

Post model creation, we move to evaluation phase. In this phase we would try to access the efficacy of the model. With generalized model approach adopted for this study, we have seen accuracy of 73% on question-answer pipeline and 68% on sentiment analysis.

In Deployment phase, the model was handed over to Auditors for them to check if the model can assist them in reducing the time taken for audits.

# Chapter 6: Business Understanding

The finance department has been relentlessly working on resolving customer's dispute related to invoice generation. Customers would approach with their concerns on Invoices through an email interface. A ticket is created against each such query and assigned to an analyst responsible for resolving the query. While doing so, analysts must ensure timely resolutions and thereby providing superior customer service. However, how do we measure quality of such responses from thousands of responses? Here comes the quality audit team, tasked to pick a random sample of such responses, and score as per the following parameters:

| Parameter | Detailed Questions |
|---|---|
| Issue Identification | Did the employee understand customer requirements? |
| Courtesy/Empathy | Did the employee display courteousness/empathy? |

Table 6.1- **Audit Parameters**

However, not all of these may be applicable for a particular audit sample. Wherever, an auditor notices any deviation from the above parameters, that field is marked down. If any of these parameters is not applicable in the current context, it is marked as "NA".

Based on these findings, an audit is scored. These findings are further shared with each analyst responsible for resolving the customer query. Analysts may contest the findings which many a times appear to be ambiguous as the interpretation by auditors and analyst may vary significantly due to subjective nature of the parameters. A lot depends how an audit has been interpreted both by auditors and analysts.

So far, it is all about how an auditor interprets an audit. Is there any alternate way of looking at this audit task? Can we remove some present barriers like interpretations issues through an alternate method? Can we aid our auditors with some information that would help them take better decision? Can we help them focus only on the critical aspect of the audit samples like where the queries are complex in nature and not very straight through? Can we help them audit more such pain areas? All these led us to explore machine learning based approach where straight through queries can be audited using per-trained NLP models.

In our study, we have observed that most these queries (approx. 80%) are usually resolved in one or two responses. Can a pre-trained model take care of these kind of samples? These are certain questions we would like to explore through this study.

# Chapter 7: Data Understanding

Data in this study is in form of text which is generated when customers approach with their queries through emails. A snippet of the mail communication is shown below:

```
Response By Email (Ma She)  (06/18/2020 05:43 AM)

Hi Angela,

I will email you when i am available , so i can set up a Zoom call for you
to process the payment.

Regards,
Ma She
Ma She | Lead Analyst | US Collections
Phone : +1-888-7414 Ext : 119
Invoice questions? Visit our website xxx.com/invoicing

Note By (Administrator)  (06/16/2020 10:46 AM)

ATTENTION!
List of additional requested contact(s) for the incident is:
naveen.jd@xxx.com
james.anders1@peps.com
madhuri.sheth@xxx.com


Customer By Service Email ()  (06/16/2020 10:46 AM)

Mri,

Good afternoon.

I have tried calling you on a couple of occasions to give you my credit
card number since you are unable to accept it via email or send me an
online payment link.  Could you advise your availability as I'd like to
connect with you to get this invoice paid.

Thank you,

Aela
```

```
   /*<![CDATA[*/ P.MsoNormal { MARGIN: 0px } LI.MsoNormal { MARGIN: 0px }
 DIV.MsoNormal { MARGIN: 0px } P.MsoListParagraph { MARGIN-LEFT: 48px }
 LI.MsoListParagraph { MARGIN-LEFT: 48px } DIV.MsoListParagraph { MARGIN-LEFT:
48px } HR.style8 { BORDER-TOP: #8c8b8b 1px solid; BORDER-BOTTOM: #fff 1px solid
 } HR.style8:after { BORDER-TOP: #8c8b8b 1px solid; BORDER-BOTTOM: #fff 1px
solid; MARGIN-TOP: 2px; DISPLAY: block } /*]]>*/

      Subject


      ABCD Invoice 6718702 Available for Review

              Contact Information

              Email Address: xxxx@xxx.com
              First Name: No Value
              Last Name: No Value

      Question Reference # 200617-000529

        Assigned: Ma She
        Date Created: 06/16/2020 10:46 AM
        Date Last Updated: 06/18/2020 05:43 AM
        Status: Closed


      Response By Email (Ma She)  (06/18/2020 05:43 AM)

      Hi Angela,

      I will email you when i am available , so i can set up a Zoom call for you
      to process the payment.

      Regards,
      Ma She
      Ma She | Lead Analyst | US Collections
      Phone : +1-888-7414 Ext : 119
      Invoice questions? Visit our website xxx.com/invoicing

      Note By (Administrator)  (06/16/2020 10:46 AM)

      ATTENTION!
```

Fig.7.1 – **Mail Communication**

It contains the following parts.

- **Question Reference** – Generated when an email is sent by customer. This is assigned to an analyst responsible for resolving the query. It is also called Incident Numbers.
- **Subject** – As specified by customer.

- **Mail from Customer** – Appears under Customer By Service Email
- **Response by analyst** – Appears under Response By Email

Similar such emails become the source of the sample data. The auditor team then follows the below sequence of steps from sampling to auditing.

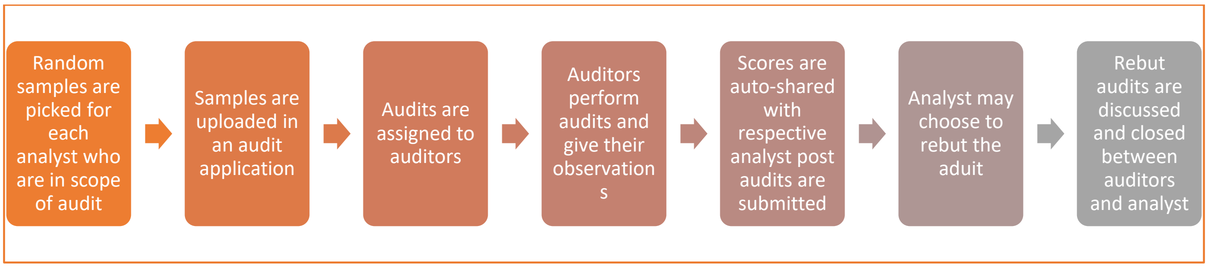


Fig.7.2 - **Audit Process**

A typical audit finding would look like this:

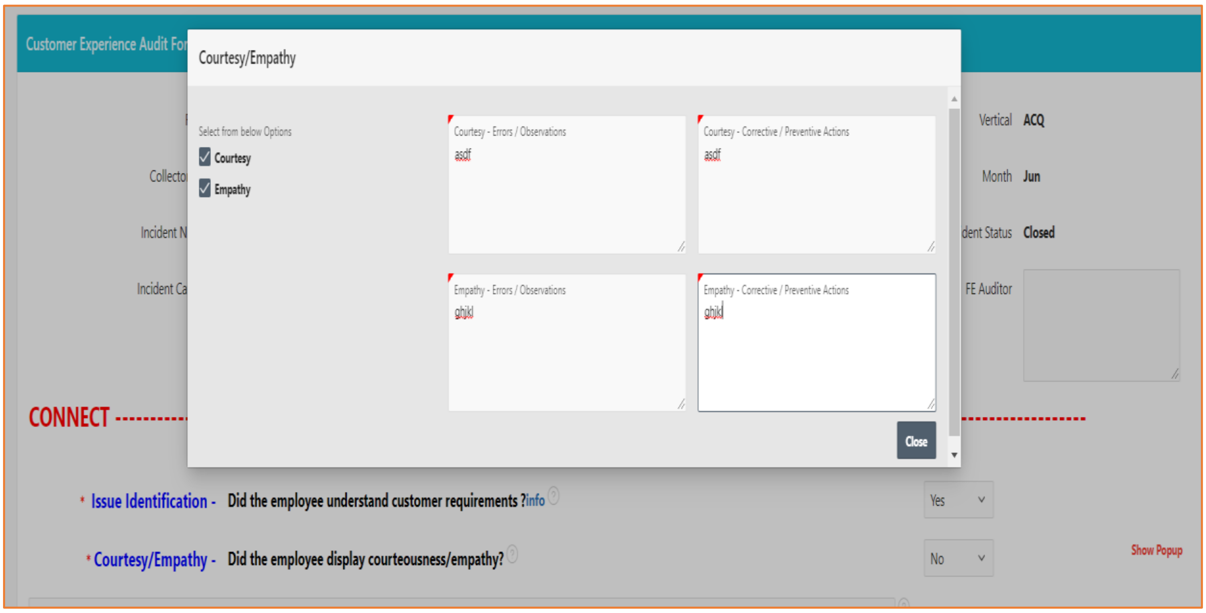


Fig.7.3 - **Audit Tool**

As shown above, the audit has been performed on 2 parameters where Courtesy / Empathy is marked down by the auditors providing justifications in the popup box.

A detailed definition of the parameters is presented below. It contains parameters under which the audits should be performed, various checkpoints and their operational definitions and the questions that need to be considered while marking audits.

| Parameter | Checkpoints | Detailed Questions | Operational Definitions |
|---|---|---|---|
| Issue Identification | • Probing<br>• Clarifying<br>• Summarizing<br>• Rephrasing<br>• Tacit Understanding | Did the employee understand customer requirements? | • **Probing:** An inquiry performed to get more information<br>• **Clarifying:** To express understanding with an explanation, examples and/or more details<br>• **Summarizing:** Provide a brief statement of the critical items<br>• **Rephrasing:** To express understanding in an alternative way<br>• **Tacit Understanding:** Identify the query or issue when probing, clarifying, summarizing, and rephrasing is not required |
| Courtesy/Empathy | • Courtesy<br>• Empathy | Did the employee display courteousness/empathy? | • **Courtesy:** Polite behavior/action (apologetic, thankfulness, respectful & considerate)<br>• **Empathy:** Ability to understand situations and how someone else feels |

Table 7.1- **Audit Checkpoints and Operation Definition**

Through this study we aim to see how many of these checkpoints can be tackled through per-trained models available as open source.

# Chapter 8: Data Preparation

Since the data used in the model is in form of text, the first step of data preparation involves extracting the responses from the mails and mapping those back to the corresponding queries. Each such sequence of to-and-fro mails involving customer queries and their corresponding responses is captured under one incidence which is unique. We have also observed a pattern in which these mails are getting recorded. The only limitation that we have encountered is getting the dump of incidents with emails. A manual approach has been followed at this moment; however, the plan is to get the mail dump by querying the database. Also, we could not use any web-scrappers to scrape the data due to security concerns. Hence took the manual of downloading each incident and saving it in text format as appearing in the Fig.1.

Post data download, we extracted the customer mails and their corresponding responses through a data extraction script written using python.

Once we have extracted the data, the next step is to correctly feed those into the model. Presently we have used the python codes to build our question-answer pipeline and sentiment analysis pipeline.

# Chapter 9: Data Modeling

Based on our problem statement and the data availability we have decided to use 'Hugginface Transformer' architecture. It provides general-purpose architectures with more than 32 pre-trained models. It also reduces the need of writing codes to just three lines. The whole purpose of this study boils down the fact that if a business problem specific to a particular organization can be tackled through these kinds of pre-trained models, how much of these can be put into use straight way. For the purpose of this study, we have decided to divide it two different parts with each part trying to address a particular business need.

We discuss our modeling approach as follows:

- **Part One (Context based response)** – It tackles the need of first parameter which is "Issue Identification" defined by the question 'Did the employee understand customer requirements?' Here, we have used 'transformers question-answer' pipeline using 2 different models ('**DistilBertForQuestionAnswering'** and '**bert-large-uncased-whole-word-masking-finetuned-squad'**) to identify if the response provided is as per the context / query received from customer. If the model can throw a score & answer, we consider it as a contextual answer otherwise we treat it as not a direct answer to the customer query and hence would require manual intervention and understanding.

- **Part Two (Sentiment Analysis)** – This part tackles the question of 'Did the employee display courteousness / empathy?' while responding to customer query. Here we have used 'transformers sentiment-analysis' pipeline using the pre-trained model '**distilbert-base-uncased-finetuned-sst-2-english**'. However, one noticeable limitation of this approach is repeated use of negative words like 'not', 'no', 'cannot', and 'unable' in one sentence can give a different result.

The data collected for the purpose is 70 different emails conversations between customer and the analyst. Each such emails are then segregated into query and its corresponding response. For the sake of simplicity, we have chosen only those queries which were answered in one response. As stated earlier, these types of responses constitute to 60% of overall incidents. This also helped us extract the response and the related query in a relatively easy manner. Those are then fed into ‘question-answer’ pipeline model as shown in the data pipeline diagram. The outcome can either have a score & answer or a ‘key error’ remark.

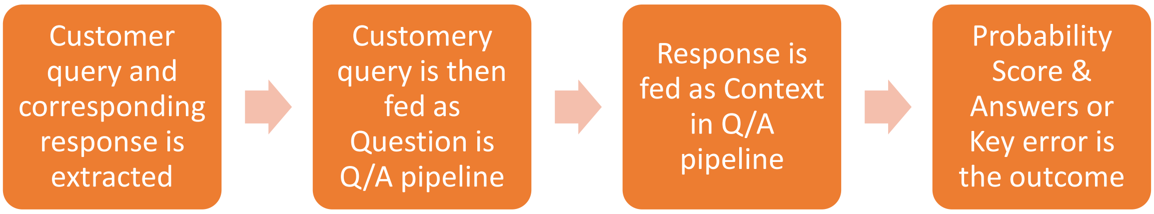


Fig.9.1 – **Data Pipeline for ‘question-answer’**

Code snippet is also provided below to demonstrate the above approach.


```
[ ] ('invoices', 'NNS'),
    ('in', 'IN'),
    ('portal', 'NN')]

from transformers import pipeline

nlp = pipeline("question-answering")

context = r"""Greetings from ABCDE! The account balance is $500 """
q = "What is my accout balance?"
try:
  print(nlp(question=q, context=context))
except:
  print("Key Error")

{'score': 0.9700969673887966, 'start': 45, 'end': 49, 'answer': '$500'}
```

Fig.9.2 – **Code Snippet ‘question-answer’**

For sentiment analysis, our approach is using transformers sentiment-analysis pipeline on responses to see if the response is positive or negative. Positive and neutral may not require any further manual analysis, however negative analysis may require further manual check.


Fig.9.3 – **Data Pipeline ‘sentiment analysis’**

Snapshot of the code is provided below to demonstrate the above approach.

```
model = AutoModelForSequenceClassification.from_pretrained("distilbert-base-uncased-finetuned-sst-2-english")

nlp = pipeline('sentiment-analysis', model='distilbert-base-uncased-finetuned-sst-2-english')

# Statement to be predicted the sentiment of
statement = "Hope you are safe & fine. Attached is the AKQA SOA for open & overdue invoices"
output = nlp(statement)
# Pipeline of 'sentiment-analysis'
print(output)
```

```
[{'label': 'POSITIVE', 'score': 0.990807831287384}]
```

Fig.9.4 – **Code Snippet ‘sentiment analysis’**

# Chapter 10: Data Evaluation

Our approach here is to test pre-trained model on a business problem and it worked fine for this study. For the sake of this study, we have chosen scenarios where the queries are resolved in one response which as whole contribute to 60% of the total query received.

Below is the snapshot of the results. As of now, we can get decent outcome from the model vis-à-vis the original result. It is a good start considering the fact that we have used a generalized pre-trained model on a business scenario.

| Model | Context | Question | Model Score | Key Error | Model Answer | Original Result | Model Performace |
|---|---|---|---|---|---|---|---|
| bert-large-uncased-whole-word-masking-finetuned-squad | We need clear information on refund that is processed, so looping in cash application team. ... | Hi there, I am the Cloud Platform Rep for Federated Co-op. I was speaking with Bra Zi, Associate VP of Tech today, and they recently received this cheque which I have attached. There is a pretty good sum of money here, but no real explanation as to what it is for. We have completed 2 different sets of approvals for Credits, but nothing of this magnitude. My thoughts are this could be for the ULA credit as they have paid for database support twice this year (2 for ULA support, 2 for UCC Credits). But if this in fact the case, are we able to provide an explanation of the dollar amount included? If this inquiry needed to go to a different group please let me know and I can talk with them about this as well. Thanks, | 0.005736943 | NA | looping in cash application team | in context | |
| bert-large-uncased-whole-word-masking-finetuned-squad | Greetings from ABCDE! Could you please share complete customer name and customer number for us to check and assist? We are unable to find any data with number mentioned in subject line | Could you please check if this customer has any outstanding balance? Thank you in advance | 0.071973279 | Yes | share complete customer name and c | in context | |
| bert-large-uncased-whole-word-masking-finetuned-squad | Greetings from ABCDE! Could you please share complete customer name and customer number for us to check and assist? We are unable to find any data with number mentioned in subject line | Could you please check if this customer has any outstanding balance? Thank you in advance | | Yes | | Not in Context | |
| bert-large-uncased-whole-word-masking-finetuned-squad | Hope you are safe & fine. Attached is the AKQA SOA for open & overdue invoices ... | Can I get the most updated SOA for AKQA including all open and closed invoices? | 0.604060338 | NA | Attached | in context | |
| bert-large-uncased-whole-word-masking-finetuned-squad | Greetings from ABCDE! The account balance is $500 | What is my accout balance? | 0.970096967 | NA | $500 | in context | |
| bert-large-uncased-whole-word-masking-finetuned-squad | As checked ABCDE is not yet on boarded to Ari portal, hence we cannot upload the invoices in portal.. | Can we upload invoices in portal? | 0.222692266 | NA | cannot upload the invoices in portal | in context | |
| bert-large-uncased-whole-word-masking-finetuned-squad | As checked ADEF is not yet on boarded to OC portal, hence we cannot upload the invoices in portal.. | ADEF asked if it is possible to post the two invoice in ariba. I have not access so I wanted to check if you would be able to do it? | 0.076252647 | NA | we cannot upload the invoices in port | in context | |

Table 10.1 – **Using bert-large-uncased-whole-word-masking-finetuned-squad**

While the Question-pipeline gave us decent results, the same result could not be achieved using Sentiment-analysis pipeline. Below is the snapshot of the results:

| Model | Context | Model Score | Model Answer | Original Result | Model Performace |
|---|---|---|---|---|---|
| distilbert-base-uncased-finetuned-sst-2-english | Greetings from ABCDE! Hope this mail finds you well. The outstanding balance is $5000 | 0.99801594 | POSITIVE | Positive | |
| distilbert-base-uncased-finetuned-sst-2-english | As checked Abcde is not yet on boarded to Ari portal, hence we cannot upload the invoices in portal | 0.99857223 | NEGATIVE | Neutral | |
| distilbert-base-uncased-finetuned-sst-2-english | Hope you are safe and healthy..!!With reference to the below email, do you want us to update the billing contact as Beverly Smith <bhmith@sfasu.edu> for all future billings and invoices. If so please confirm, so that we can raise the request to update the records. Please do let me know, if you have any questions. I will be happy to help.'" | 0.996198773 | POSITIVE | Positive | |
| distilbert-base-uncased-finetuned-sst-2-english | Hope you are safe & fine. Attached is the AKQA SOA for open & overdue invoices ... | 0.991036952 | POSITIVE | Positive | |
| distilbert-base-uncased-finetuned-sst-2-english | We need clear information on refund that is processed, so looping in cash application team. ... | 0.998514116 | NEGATIVE | Negative | |
| distilbert-base-uncased-finetuned-sst-2-english | Could you please share complete customer name and customer number for us to check and assist? | 0.952159643 | NEGATIVE | Positive | |
| distilbert-base-uncased-finetuned-sst-2-english | The Dispute has been raised from Collections end. | 0.966411412 | NEGATIVE | Neutral | |
| distilbert-base-uncased-finetuned-sst-2-english | Please let us know, are your looking for the invoice or the payment method! Accordingly we would do our best to resolve your requirements. | 0.656778991 | POSITIVE | Positive | |
| distilbert-base-uncased-finetuned-sst-2-english | Hope you are doing well!As mentioned by Nab looking forward to work with you. We look forward for the short payment of the invoice. | 0.999776661 | POSITIVE | Positive | |

Table 10.2 – **Using distilbert-base-uncased-finetuned-sst-2-english**

# Chapter 11: Deployment

The model was handed over to Auditors for them to test its applicability. Auditors would feed the model with context and responses and use the result to fine tune their findings. Initial trials show encouraging results. The model was helping auditors save time on auditing and reducing manual interpretation errors to a great extent. Post the trial phase, which we plan to continue for another month or two, we plan to deploy the model within a particular line of business.

# Chapter 12: Analysis and Results

The model was tested on the actual audit results and below is the summary of model performance.

| Model Used | Sample Tested | Type | Model Outcome | Interpretation | Model Accuracy on actual Data | Business Recommendation |
|---|---|---|---|---|---|---|
| bert-large-uncased-whole-word-masking-finetuned-squad | 70 | Question-Answer | Probability score & Answer | The answer to customer query is in line with the context( customer query) where probability score is > than 0.005 | **73%** | As long as the Outcome has a probability score > 0.005 with answer, we are fine with the outcome, else manual check is recommended |
| | | | Key Error | The answer to customer query is either not in line with the context( customer query) or the model is not able to understand the context | | Key Error can be hits or misses, hence these require manual verification |
| distilbert-base-uncased-finetuned-sst-2-english | 70 | Sentiment | Positive or Negative | As long as the outcome is 'Positive', we interpreted it in line with the required tone whereas if the outcome is 'Negative' we cannot conclude it as a negative tone. | **68%** | Positive result is fine whereas Negative outcome needs manual verification as presence of negative words like "Unable, not, no, don't" tend to produce "Negative" outcome by the model which may be case always |

Table 11.1 – **Model Outcome**

A sample size of 70 incidents where there is only one response are taken and validated with the actual findings. The Question-Answer pipeline has an accuracy of 73% whereas Sentiment Analysis has 68%.

Based on these findings, we tried to quantify the benefits as shown below.

| Audit Type | Samples Audit (a) | Avg. Time (in Min) to read & interpret each audit (b) | Total Time Taken (c)=a*b | Interpretation Error / Model Error (d) | Rework Required on samples (e) = c*d | Estimated Time (in Min) on rework (f) = e*10 | Total time (in Min) (g)=c+f |
|---|---|---|---|---|---|---|---|
| Manual | 70 | 10 | 700 | 12% | 8.4 | 84 | 784 |
| Using Model | 70 | 1 | 70 | 29% | 20.3 | 203 | 273 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Estimated time (in Min) saved per Audit (h) = g / a | 7.3 | Avg. number of Audits per month (i) | 287 | Total Time Saved (in Min) (j) = h*i | 2,095 | Total Time Saved (in hrs.) k = j / 60 | 35 | FTE Benefit per day k / (6.5*18) | **0.30** |

Table 11.2 – **Benefit Analysis**

We are close to achieving one third of FTE benefit per day considering 18 working days per month with 6.5 of effective work hour per day. This is tremendous feet since we have used a pre-trained generalized model for solving a business problem.

However, with the above benefits, we do have a few drawbacks in our model. To understand it better we conducted a SWOT analysis of our approach. The same is presented below.

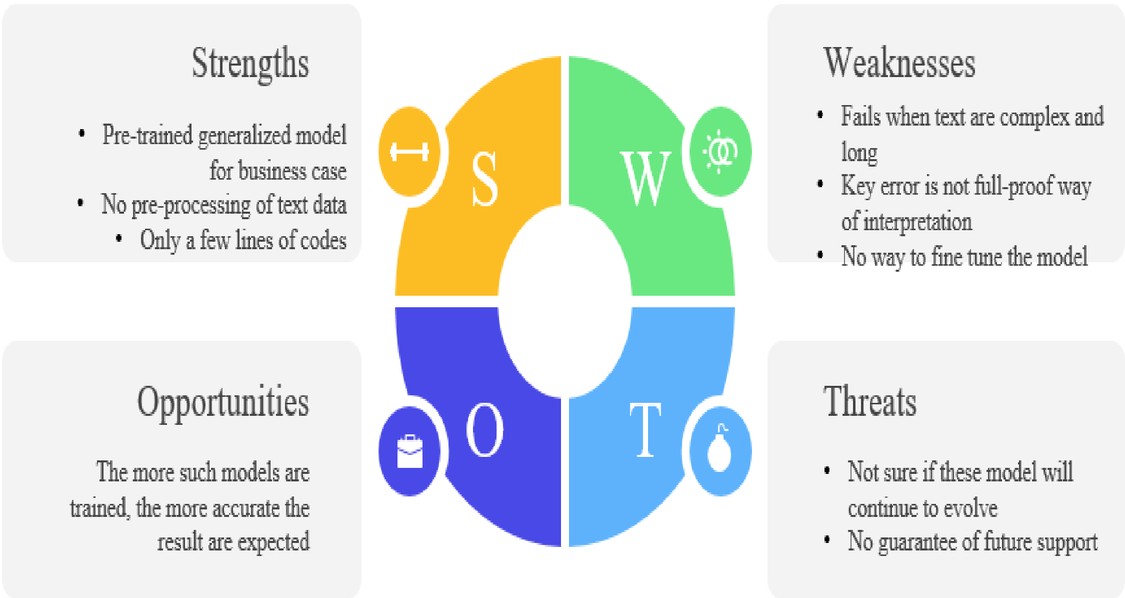


Fig.11.1 – **SWOT Analysis**

# Chapter 13: Conclusions and Recommendations for future work

There was problem at hand – a business problem specific to a particular organization. And now we have a solution built only using available per-trained model architectures. We can address almost 60% of our requirement. This proves that these state-of-the-art open architectures available as free source have immense potential in future. Future versions of these models may be able to address most of real-world business requirements. All this at no extra cost to a company.

Yes, we have seen limitations of this kind of approach while solving a business problem as many of these may be specific to a particular organizations / industry and complex in nature. We can always invest in building our own machine learning models to address these complex scenarios; however, our approach is extremely useful where a particular department within the organization may not have the capability to invest and build a separate analytics wing.

This approach paves way for quick adoption of state-of the art open-source natural language pre-trained models, without reinventing the wheel, for addressing low-hanging fruits which may not be overly complex in nature.

In conclusion, what we have seen is unstructured data like text earlier involved a lot of pre-processing which is the most time-consuming part of any NLP projects, but with the advent of pipeline approach, three lines of code is able to do everything starting from preprocessing to final outcome. This is simply amazing and super time-saver. As for any project, time and cost often overshoots the budget, whereas these kinds of architectures solve all these problems.
We expect to explore some more of these architectures and explore ways to solve business problems.  However, if we are looking for high quality output, it is always advisable to build our own data repository / lake, create our own models, train the models using business specific data and deploy it to solve a specific need.

# Bibliography

*A Visual Guide to Using BERT for the First Time – Jay Alammar – Visualizing machine learning one concept at a time.* (n.d.). Retrieved September 13, 2020, from http://jalammar.github.io/a-visual-guide-to-using-bert-for-the-first-time/

Aken, B. van, Winter, B., Löser, A., & Gers, F. A. (2020). *VisBERT: Hidden-State Visualizations for Transformers*. 207–211. https://doi.org/10.1145/3366424.3383542

Jobanputra, M. (2019). *Question Answering for Fact-Checking*. 52–56. https://doi.org/10.18653/v1/d19-6609

M'Haimdat, O. (n.d.). *Sentiment Analysis iOS Application Using Hugging Face's Transformers Library | by Omar M'Haimdat | Heartbeat*. Retrieved September 13, 2020, from https://heartbeat.fritz.ai/sentiment-analysis-ios-application-using-hugging-faces-transformers-library-7ae4df1b3150

*Quick tour — transformers 3.1.0 documentation*. (n.d.). Retrieved September 13, 2020, from https://huggingface.co/transformers/quicktour.html

*Rule-based matching · spaCy Usage Documentation*. (n.d.). Retrieved September 14, 2020, from https://spacy.io/usage/rule-based-matching

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. 2–6. http://arxiv.org/abs/1910.01108

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to Fine-Tune BERT for Text Classification? *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 11856 LNAI*(2), 194–206. https://doi.org/10.1007/978-3-030-32381-3_16

van Aken, B., Winter, B., Löser, A., & Gers, F. A. (2019). *How Does BERT Answer Questions?* 1823–1832. https://doi.org/10.1145/3357384.3358028

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *2017-Decem*(Nips), 5999–6009.

Wiegreffe, S., & Pinter, Y. (2020). Attention is not not explanation. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference*, 11–20. https://doi.org/10.18653/v1/d19-1002

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. Le, Gugger, S., … Rush, A. M. (2019). *HuggingFace's Transformers: State-of-the-art Natural Language Processing*. http://arxiv.org/abs/1910.03771

*Working with Hugging Face Transformers and TF 2.0 | by Akash Desarda | Towards Data Science*. (n.d.). Retrieved September 13, 2020, from https://towardsdatascience.com/working-with-hugging-face-transformers-and-tf-2-0-89bf35e3555a

*A Visual Guide to Using BERT for the First Time – Jay Alammar – Visualizing machine learning one concept at a time.* (n.d.). Retrieved September 13, 2020, from http://jalammar.github.io/a-visual-guide-to-using-bert-for-the-first-time/

Aken, B. van, Winter, B., Löser, A., & Gers, F. A. (2020). *VisBERT: Hidden-State Visualizations for Transformers*. 207–211. https://doi.org/10.1145/3366424.3383542

Jobanputra, M. (2019). *Question Answering for Fact-Checking*. 52–56. https://doi.org/10.18653/v1/d19-6609

M'Haimdat, O. (n.d.). *Sentiment Analysis iOS Application Using Hugging Face's Transformers Library | by Omar M'Haimdat | Heartbeat*. Retrieved September 13, 2020, from https://heartbeat.fritz.ai/sentiment-analysis-ios-application-using-hugging-faces-transformers-library-7ae4df1b3150

*Quick tour — transformers 3.1.0 documentation*. (n.d.). Retrieved September 13, 2020, from https://huggingface.co/transformers/quicktour.html

*Rule-based matching · spaCy Usage Documentation*. (n.d.). Retrieved September 14, 2020, from https://spacy.io/usage/rule-based-matching

Sanh, V., Debut, L., Chaumond, J., & Wolf, T. (2019). *DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter*. 2–6. http://arxiv.org/abs/1910.01108

Sun, C., Qiu, X., Xu, Y., & Huang, X. (2019). How to Fine-Tune BERT for Text Classification? *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11856 LNAI*(2), 194–206. https://doi.org/10.1007/978-3-030-32381-3_16

van Aken, B., Winter, B., Löser, A., & Gers, F. A. (2019). *How Does BERT Answer Questions?* 1823–1832. https://doi.org/10.1145/3357384.3358028

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *2017-Decem*(Nips), 5999–6009.

Wiegreffe, S., & Pinter, Y. (2020). Attention is not not explanation. *EMNLP-IJCNLP 2019 - 2019 Conference on Empirical Methods in Natural Language Processing and 9th International Joint Conference on Natural Language Processing, Proceedings of the Conference,* 11–20. https://doi.org/10.18653/v1/d19-1002

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Scao, T. Le, Gugger, S., … Rush, A. M. (2019). *HuggingFace's Transformers: State-of-the-art Natural Language Processing*. http://arxiv.org/abs/1910.03771

*Working with Hugging Face Transformers and TF 2.0 | by Akash Desarda | Towards Data Science*. (n.d.). Retrieved September 13, 2020, from https://towardsdatascience.com/working-with-hugging-face-transformers-and-tf-2-0-89bf35e3555a

## Appendix

### Plagiarism Report[1]

Automating Customer Experience Audit using Per-trained Generalized Models

ORIGINALITY REPORT

| 7% | 6% | 1% | 5% |
|---|---|---|---|
| SIMILARITY INDEX | INTERNET SOURCES | PUBLICATIONS | STUDENT PAPERS |

PRIMARY SOURCES

| | | |
|---|---|---|
| 1 | global.oup.com<br>Internet Source | 2% |
| 2 | Submitted to Sogang University<br>Student Paper | 1% |
| 3 | export.arxiv.org<br>Internet Source | <1% |
| 4 | towardsdatascience.com<br>Internet Source | <1% |
| 5 | huggingface.co<br>Internet Source | <1% |
| 6 | pdf.aminer.org<br>Internet Source | <1% |
| 7 | jalammar.github.io<br>Internet Source | <1% |
| 8 | Submitted to Regional Centre for Biotechnology<br>Student Paper | <1% |
| | Submitted to Lovely Professional University | |

### Publications in a Journal/Conference Presented/White Paper[2]

### Any Additional Details

[1] Turnitn report to be attached from the University.

[2] URL of the white paper/Paper published in a Journal/Paper presented in a Conference/Certificates to be provided.