REVA UNIVERSITY
Bengaluru, India

A Project Report on

Automated Literature Review with Transformers

Submitted in Partial Fulfilment for Award of Degree of

Master of Business Administration

In Business Analytics

Submitted By

Sanjeev Kumar Jha

R19MBA06

Under the Guidance of

Dr. J. B. Simha

CTO, Abiba Technologies/Chief Mentor, RACE

REVA Academy for Corporate Excellence - RACE

REVA University

Rukmini Knowledge Park, Kattigenahalli, Yelahanka, Bengaluru - 560 064

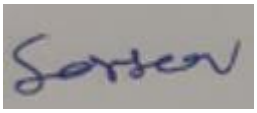
race.reva.edu.in

August, 2022

REVA UNIVERSITY
Bengaluru, India

Candidate's Declaration

I, **Sanjeev Kumar Jha** hereby declare that I have completed the project work towards the second year of **Master of Business Administration in Business Analytics** at, REVA University on the topic entitled **Automated Literature Review with Transformers** under the supervision of **Dr. J. B. Simha**, Chief Mentor. This report embodies the original work done by me in partial fulfilment of the requirements for the award of degree for the academic year 2022.

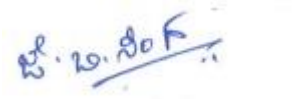


Place: Bengaluru
Date: 27th August 2022

Name of the Student: Sanjeev Kumar Jha
Signature of Student

Certificate

This is to Certify that the project work entitled **Automated Literature Review with Transformers** carried out by **Sanjeev Kumar Jha** with R19MBA06**,** is a bonafide student of REVA University, is submitting the second-year project report in fulfilment for the award of Master of Business Administration in Business Analytics during the academic year 2022. The Project report has been tested for plagiarism and has passed the plagiarism test with a similarity score of less than 15%. The project report has been approved as it satisfies the academic requirements in respect of the Automated Literature Review with Transformers work prescribed for the said degree.



Signature of the Guide
Dr. J. B. Simha
Guide

Signature of the Director
Dr. Shinu Abhi
Director

External Viva
Names of the Examiners

1. Vaibhav Sahu, Strategic Cloud Engineer, Google
2. Abhishek Sinha, Data Science Manager, Capgemini

Place: Bengaluru
Date: 27th August 2022

REVA UNIVERSITY
Bengaluru, India

Acknowledgement

I would like to thank my mentor Dr. J. B. Simha and to the director of RACE Dr. Shinu Abhi who gave me a golden opportunity to work on this project. I'd also like to express my gratitude to Prof. Mithun Dolthody Jayaprakash and Dr. Rashmi Agarwal wholeheartedly.

I must also thank my parents and friends for the immense support and help during this project. Without their help, completing this project would have been very difficult.

I would like to acknowledge the support provided by the founder and Hon'ble Chancellor, Dr. P Shayma Raju, Pro-Vice Chancellor, Dr. M. Dhanamjaya, and Registrar, Dr. N. Ramesh as a standard protocol.

Place: Bengaluru

Date: 27th August 2022

REVA UNIVERSITY
Bengaluru, India

Similarity Index Report

This is to certify that this project report titled **Automated Literature Review with Transformers** was scanned for similarity detection. Process and outcome is given below.

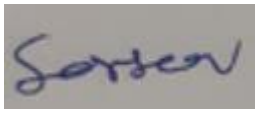
Software Used: **Turnitin**

Date of Report Generation: **26th Aug 2022**

Similarity Index in %: **13%**

Total word count: **5787**

Name of the Guide: **Dr. J. B. Simha**



Place: Bengaluru

Date: 27th August 2022

Verified by: M N Dincy Dechamma

Name of the Student: Sanjeev Kumar Jha

Signature of Student

Signature

Dr. Shinu Abhi,

Director, Corporate Training

List of Abbreviations

Sl. No	Abbreviation	Long Form
1	NLP	Natural Language Processing
2	TM	Text Mining
3	SLRs	Systematic Literature Reviews
4	LSA	Latent Semantic Analysis
5	BERT	Bidirectional Encoder Representations from Transformers
6	GPT-2	Generative Pretrained Transformer 2
7	Rough	Recall-Oriented Understudy for Gisting Evaluation
8	TF-IDF	Term Frequency-Inverse Document Frequency
9	RNN	Recurrent Neural Network

List of Figures

List of Tables

Abstract

Business research is producing new knowledge at an astounding rate, yet it is still fragmented and heterogeneous. This makes it challenging to stay up to date, be at the forefront of research, and interpret the data in context in a particular area of business research. Because of this, it is more crucial than ever as a research strategy. Traditional literature reviews usually lack thoroughness and rigour and are conducted haphazardly rather than in accordance with a predetermined methodology. As a result, questions about the validity and reliability of these kinds of evaluations may be brought up. Despite the value of conducting Systematic Literature Reviews (SLR) of the literature to pinpoint research needs across various disciplines of study, manually doing SLR is a challenging, multi-stage, and time-consuming process.

The primary goal of this research topic is to extract the reference paper and summarize it based on existing research papers in the same area and generate a Summary of the text which prevents duplication.

The text has been taken from different papers and after that with the help of different regular expressions collected data was cleaned from unnecessary words or punctuation marks. On the cleaned text different NLP techniques (Text Rank, Lex Rank, LSA, Luhn, KL Sum, BERT, GPT-2) have been utilized for summarization of the text. Out of all NLP techniques it has been observed that GPT and BERT were giving the best result.

On a selected topic, different research papers have been explored. Literature reviews of different research papers have been taken and summarized for reference to new researchers in that field. In this way, this model reduces the research time of different researchers and gives them an idea of previous research which has taken place on this topic in recent years.

Keywords: Text Summarization, Transformer, Research Paper, Extractive Summarization, Transformer, NLP

Contents

Chapter 1: Introduction

A literature review looks at information that has been published on a particular subject and, on occasion, information from a certain time period in a particular field. A mature subject addresses the need for critique of the topic's expanding and more varied body of knowledge as well as prospective reconceptualization. The second kind of it is centred on cutting-edge or unique topics that could profit from a thorough conceptualization and synthesis of the literature. Because of the novelty of these topics and the lack of a comprehensive literature review, A new model or framework, as well as a first or preliminary conception of the subject, are more likely to emerge from the review.

There are several potential justifications for doing this, including gathering data for the creation of evidence-based care, both as a step in the research process and as a component of an academic assessment. Conducting it seems to be a challenging task to too many capable students. The most often queries range from what a literature review comprises to how to choose a subject, how many articles to include, and where to start (Cronin, Ryan, & Coughlan, 2008).

1.1 Literature Review Understanding:

In this sense, it is a scientific publication that presents up-to-date information, including significant findings as well as theoretical and methodological advancements in a certain field. Secondary sources, do not report on new or unusual experimental work, as in literature reviews. These reviews which are most frequently associated with academic literature, are published in scholarly publications and should not be confused with book reviews, it might be published in the same journal. Evaluations of the literature form the basis for study in almost every academic field.

It is crucial to think of knowledge as having **three layers** while discussing a particular subject.

1. The researchers' initial studies, known as primary studies, are carried out and published.
2. Second, assessments of those results often go beyond the original investigations while summarising and proposing new interpretations based on them.
3. Third, there are unofficial judgments, opinions, and interpretations that get incorporated into the legend around the subject.

It is important to note that this third layer of knowledge is commonly referred to as "true" while writing this, despite just having a tenuous connection to the primary studies and secondary literature evaluations.

The three categories are **evaluative**, **exploratory**, and **instrumental**. The systematic review, a fourth form that is sometimes categorised independently, is simply a literature review that is concentrated on a particular research issue and makes an effort to find, evaluate, select, and synthesise all high-quality research data and arguments pertinent to that topic.
A meta-analysis is a systematic review that makes good use of statistical tools to combine data from all chosen studies in order to get a more accurate conclusion. An interactives' objective is to create new knowledge about a subject by analysing, evaluating, and synthesising the literature in question (Kennedy & M, 2017).

1.2 Types of Literature Review

The numerous types of literature reviews are listed below:

1.2.1 Argumentative Review

This approach examines the literature only when it is necessary to support or refute an established thesis, a long-held belief, or a philosophical issue. The main objective is to produce a corpus of writing that could serve as a counterargument. Considering how some social science studies are value-laden. For example, immigration regulation and educational reform. Argumentative methods of literary analysis may be appropriate and worthwhile. However, they can cause bias issues when used to generate summary claims like those seen in systematic reviews (Kennedy & M, 2017).

1.2.2 Integrative Review

Known as a study style that incorporates, analyses, evaluates, and synthesises representative literature on a subject to create fresh frameworks and viewpoints on the subject. Within the body of literature are all the research that specifically address related or parallel hypotheses. A well-done integrative review satisfies the same standards as primary research in terms of clarity, rigour, and replication (Kennedy & M, 2017).

1.2.3 Historical Review

Few things stand apart from historical precedent. Examining studies through time, historical reviews frequently start with the first time an issue, concept, theory, or phenomenon appears in the literature and trace its development within a discipline's scholarship. The objective is to contextualise research within a historical framework to demonstrate knowledge through cutting-edge discoveries and to identify potential future research fields (Kennedy & M, 2017).

1.2.4 Methodological Review

The focus of a review is often on the method of analysis rather than the content of what was said. This method offers a framework of comprehension at different levels (theory, substantive fields, research approaches, and data collection and analysis techniques), allowing researchers to draw on a variety of knowledge ranging from the conceptual level to practical documents for use in the fieldwork in many fields of ontological and epistemological consideration as well as for quantitative and qualitative analysis (Kennedy & M, 2017).

1.2.5 Systematic Review

In order to find and critically evaluate relevant research, as well as to gather, present, and analyse data from the studies included in the review, this form includes an overview of the currently available evidence pertinent to a clearly defined research question. The main focus is typically on a relatively specific empirical topic, which is frequently presented in a cause-and-effect structure, such as "How much does A contribute to B?" (Kennedy & M, 2017).

1.2.6 Theoretical Review

This form's objective is to investigate the body of theory that has been accumulated in regard to a certain subject, idea, theory, or phenomenon in a concrete way. Theoretical analysis aids in identifying current theories, their connections, the depth to which concepts have been researched, and the creation of new hypotheses to be tested. This method is widely used to show that there are no valid theories or that existing theories are insufficient to explain novel or emerging research problems. A single theoretical idea, a whole theory(Kennedy & M, 2017).

As there are total of six types of literature reviews i.e., Argumentative, Integrative, Historical, Methodological, systematic, and theoretical. Different research falls into a different category by summarizing it researcher can come to know in which category a particular paper belongs to.

Chapter 2: Literature Review

The writing of a thesis, dissertation, or journal article, among other graduate and post-graduate student assignments, typically calls for it. A study proposal or prospectus frequently includes a review of the literature (the agreement signed before a researcher officially starts writing a dissertation or thesis). It may simply be a summary of relevant sources. The key points of the source are summarised in a summary, whereas the information is reorganised or rearranged in a synthesis to help with how to analyse a research paper.

This project creates a list of previously published articles on a particular subject. A full academic work or a portion of one. In any case, the goal of this is to provide the researcher/author and the audience with a broad overview of the material that is currently accessible on the topic at hand. A thorough review of the literature helps ensure that an appropriate theoretical framework and/or research methods were used, as well as that a valid study issue was addressed. In other words, this gives the reader a perspective by setting the current work within the context of the pertinent literature. In these circumstances, the review typically comes before the sections discussing the work's procedures and results.

A review article can be included as a literature review. It is a scientific journal that delivers up-to-date information, together with significant findings and theoretical and methodological contributions to a particular field. Literature reviews are an example of a secondary source that does not present new or original experimental work. Academic publications sometimes contain such reviews, which are frequently associated with academic literature. Practically every academic discipline's research starts with a literature review.

Although some of the earliest reports mentioned in the literature were vocal, written reports make up the vast majority of reports. scholarly work that may be methodological, theoretical, empirical, or critical in nature. Second, it makes an effort to summarise, analyse, evaluate, make clear, and/or include the content of original reports(Cooper, 1988) A complete summary and critical analysis of the existing research and non-research literature on the subject at hand constitutes a literature review (Hart, 1998) (Cronin, Ryan, & Coughlan, 2008). This serves as a foundation for another goal, such providing evidence for further research into the field, while also keeping the reader informed about recent publications on the subject. A good literature

review gathers data from several sources on a particular subject. It should have a distinct search and selection process (Carnwell & Daly, 2001); (Cronin, Ryan, & Coughlan, 2008).

It differs from a report on scholarly research. Establishing a novel argument is one of the main objectives of an academic research paper, and it might be one of its components. In a research paper, the researcher used the literature to provide the groundwork and provide evidence for a novel insight. Contrarily, this aims to summarise and synthesise the thoughts and arguments of others without adding anything new.

There are several uses for this, and the vast majority of them are found in a primary research article that provides the theoretical framework for the article's main investigation. It provides the framework for the remaining parts of an academic piece. It clearly communicates to the reader the importance of earlier work while explaining the nature and content of the current body of knowledge(Okoli & Schabram, 2010).The review must synthesise the available information and provide a scientific critique of the theory in order to contribute to the work as an academic paper. It cannot simply restate the subject matter (Okoli & Schabram, 2010).

2.1 Potential algorithms for Generating literature review

Seven algorithms have been identified as having the potential to Generate it. **These are as follows:**

Text Rank: It is a graph-based text processing ranking model that demonstrates how this model may be used effectively in natural language applications. When utilising an unsupervised method to extract phrases and keywords, it is really helpful. An extractive summarization method is TextRank. It is predicated on the idea that words with higher frequency have greater significance(Barrios, L´opez, Argerich, & Wachenchauzer, 2016) (Mallick, Das, Dutta, Das, & Sarkar, 2019).

1. An extractive summarization method is TextRank.
2. It is predicated on the idea that words with higher frequency have greater significance.
3. Therefore, it is crucial to pay attention to sentences with many common words.
4. The system awards grades to each sentence in the text based on this.

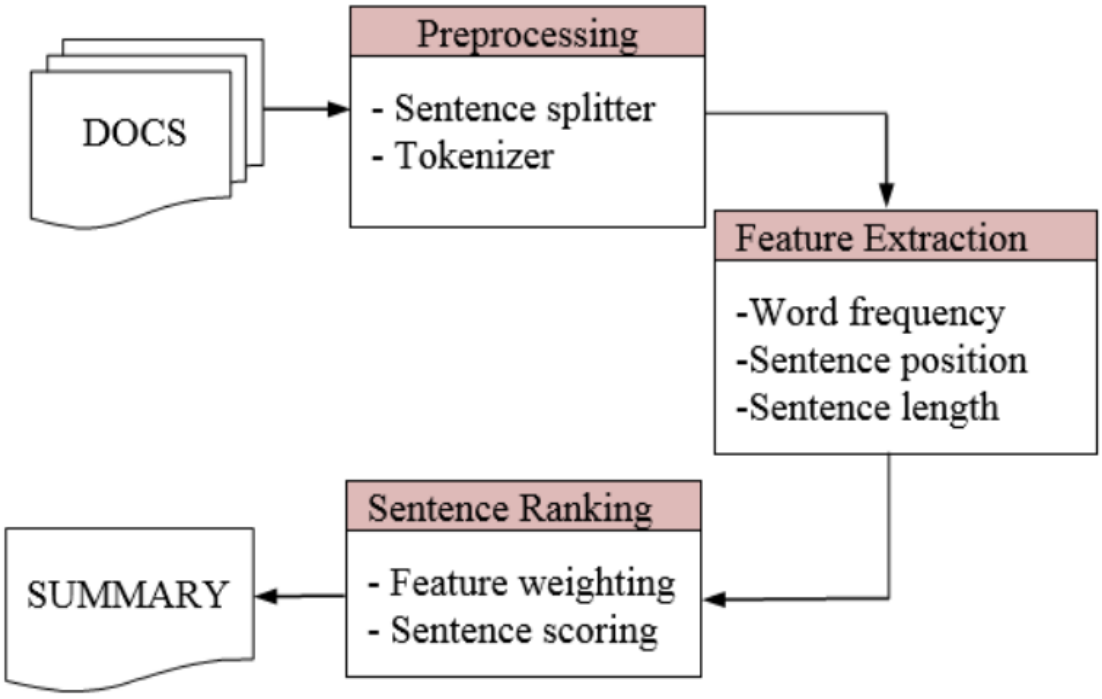


Figure No. 2.1: Text Rank Architecture for summarization

In Figure No. 2.1, text rank Architecture for summarization based on frequently, words/sentences that occurred.

Lex Rank: It is a different PageRank method offspring with the text summarising tool TextRank as a sibling. A sentence is likely to be considered important if it can be related to several other sentences in the text. Finding a sentence that is supported by other, comparable sentences and is scored higher is how LexRank works. The higher the rank of the sentence is, the higher the priority of being included within the summarized text (Radev & R., 2004) (Uçkan & Karcı, 2020).

1. An unsupervised learning approach called latent semantic analysis can be applied to extractive text summarization.
2. By applying Singular Value Decomposition (SVD) to the matrix of term-document frequency, it extracts phrases that are semantically significant.

Latent Semantic Analysis (LSA): it is a technique for examining the connections between a collection of documents and the terms they include by creating a collection of concepts pertaining to both the documents and the terms. By using SVD on the grid of term-document frequency, it extracts phrases that are semantically significant. Latent Semantic Analysis is a

generally applicable unsupervised learning approach for extractive text summarization(Ozsoy, Alpaslan, & Cicekli, 2011) (Uçkan & Karcı, 2020).

1. A sentence has a high likelihood of being significant if it is comparable to numerous other sentences in the text.
2. According to LexRank's methodology, a sentence is ranked higher if other sentences that are similar to it recommend it.
3. The priority of being included in the summary text increases with rank.

Luhn: Based on **Term Frequency-Inverse Document Frequency (TF-IDF)** method for summarising Luhn When both very low-frequency words and highly common words (stopwords), approach is quite helpful. It is used to separate sentence ranks, and the top-ranking sentences are featured in the summary(Uçkan & Karcı, 2020).

1. The technique of the Luhn Summarization algorithm is based on TF-IDF (Term Frequency-Inverse Document Frequency). When both highly frequent words (stopwords) and very low frequency words (VLF) are insignificant, it is helpful.
2. Accordingly, sentence scoring is done, and the best-scoring sentences are included in the summary.

KL Sum: It is an extractive strategy that chooses phrases based on how frequently words appear in the original text. It seeks to reduce the KL-divergence threshold (learn more). To reduce the KL-divergence, it employs a greedy optimization strategy and keeps adding sentences(Uçkan & Karcı, 2020)

1. The KL-Sum algorithm is another extraction technique.
2. Based on how closely the word distribution resembles the original text, it chooses which sentences to use.
3. It seeks to reduce the KL-divergence threshold (learn more).
4. It employs a greedy optimization strategy and continues to add phrases as long as the KL-divergence remains low.

Bidirectional Encoder Representations from Transformers (BERT): BERT is a transformer that is used to get around the long-term dependencies that Recurrent Neural Networks (RNN) and other neural networks have. It is a pre-trained naturally bidirectional model. This pre-trained model can easily complete the given NLP tasks because it is highly adaptable, which is in our case summarization (Liu, 2019) (Miller & Atlanta, 2019) (Tan & Kieuvongngam, 2020).

BERT is a transformer that is used to get around the long-term dependencies that RNN and other neural networks have. It is a pre-trained model that is naturally bidirectional. This pre-trained model can be adjusted to easily carry out the specified NLP tasks, Summarization in our case.

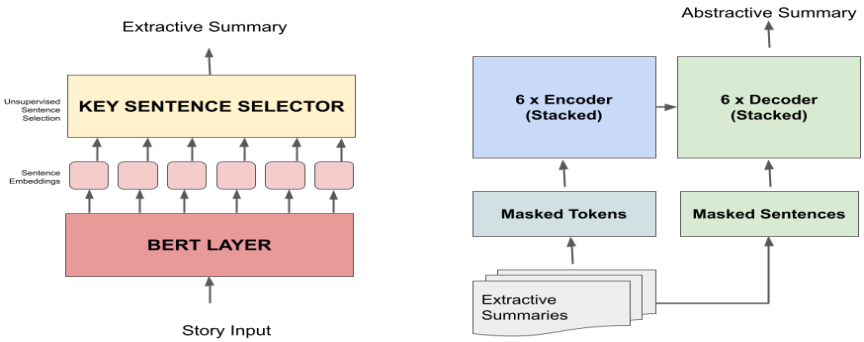


Figure No. 2.2: Bert Architecture for summarization

In Figure No. 2.2, Bert Architecture for summarization based on transformer method.

Generative Pretrained Transformer 2 (GPT-2): GPT-2 is a seq2seq model that can be altered to carry out the task of text summarization. Here the format of data is very similar to the translation task- "*text = summary*" (Tan & Kieuvongngam, 2020) (Khandelwal, Clark, Jurafsky, & Kaiser, 2019).

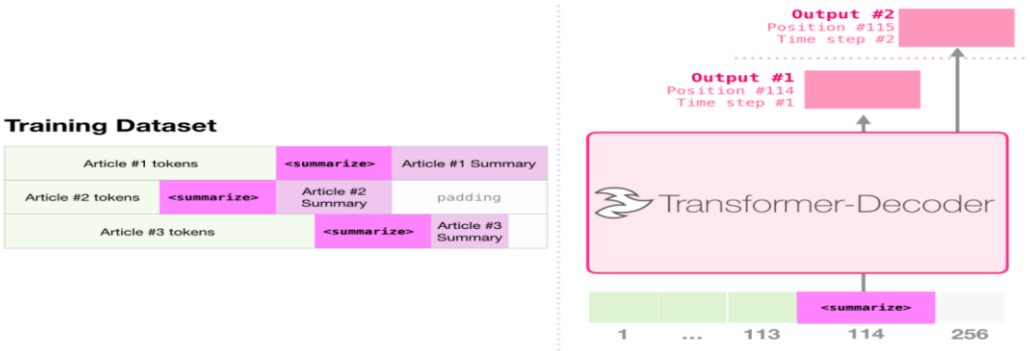


Figure No. 2.3: Gpt-2 Architecture for summarization

In Figure No. 2.3, Gpt-2 Architecture for summarization based on transformer method.

2.2 Potential matrices to evaluate the score

Recall-Oriented Understudy for Gisting Evaluation-N (Rough-N): Between the model-generated text and references RAUGE-N measures the number of matching n-grams. ROUGE score can be used for scoring (ShafieiBavani, Ebrahimi, Wong, & Chen, 2017) (Ganesan, 2018).

Bert Score: it is an automatic evaluation metric for text generation from existing text. The main advantage of using this metric is to calculate a score of comparable ness between each token in the candidate sentence and each token in the reference sentence (Zhang, Kishore, Wu, Weinberger, & Artzi, 2020).

When the researcher searches the literature review in different research papers it's not an easy task to understand the context of it. Usually, it takes a considerable time to read and understand it fully. It is quite difficult to summarize the whole literature review in a few words manually by a research scholar. When they need to refer to a greater number of papers on any topic it becomes a very time-consuming and tough task. So, there was a need for a method or model which can resolve this issue and provide some relief to the research scholars.

The text summarization model takes a lot of text for processing that’s why it needs a high-end configurations system with high computation intensive CPU\GPU otherwise for the normal system it’s a time-consuming task.

Chapter 3: Problem Statement

There are quite a few research papers are being published on daily basis. It is difficult to read all papers to start research. Also, the literature reviews are not user-friendly to be searched, causing the search complexity of the paper. This paper is focused to address this problem statement. On the other hand, a mass number of research papers usually take a decent amount of time to read across and infer into their paper.

This project could help the researchers by saving their time and energy in understanding different papers' contexts by reading abstracts and/or Literature reviews. It could give the whole context in new words and that could be understood by the researcher easily. Hence, they can use their effort in their research instead of just reading the existing work done by another researcher.

Chapter 4: Objective of Study

The objectives of this study are to summarize the research paper using different papers on the same topic. On a selected topic, different research papers have been explored. Literature reviews of different research papers have been taken and summarized for reference to new researchers in that field. Using this summarization technique, the researchers could save time in writing the literature review, as literature reviews along with paper references are used to generate by the NLP summarization technique.

Details of the objective are below: -

4.1 Choosing a review subject

Choosing a topic is the first and frequently hardest part of writing a literature review. (Timmins & McCabe, 2005); (Cronin, Ryan, & Coughlan, 2008). Lack of expertise in the topic area frequently makes the procedure harder. Here are some suggestions to assist you in making a decision more quickly.

5.1.1: Discover the general topics of interest in the field by reading the textbook.

5.1.2: To better comprehend the language (essential words), primary investigators, and issues or disputes in the area, read the chapters pertaining to the subjects that were selected.

5.1.3: Knowing the parts of a subject the reviewer is interested in as well as how much information is accessible on the subject can be determined by talking with others, such as specialists, or reading about it. (Timmins & McCabe, 2005).

4.2 Finding and selecting relevant articles

The next stage after choosing a topic is to conduct an orderly search for pertinent and related information. It is believed that a methodical approach is more likely to produce a review that could help guide practise (Cronin, Ryan, & Coughlan, 2008). Reviewers should look for thoroughness and relevance; also, the more detailed the topic or inquiry, the more focused the result (Cronin, Ryan, & Coughlan, 2008).

4.3 Analysing and synthesizing the research

After gathering the articles, they intend to employ in their research, researchers are prepared to evaluate each article (dissect it and locate the relevant information within it) and then

synthesise the collection of articles (integrate them and identify the conclusions that can be drawn from the articles as a group)(Cronin, Ryan, & Coughlan, 2008)(Sally, 2013).

4.4 Writing the review in a systematic manner

Its main objective is to aid the researcher in understanding the kind of research paper they think is required. The secret to a good research report is the ability to explain the findings in a way that clearly and consistently demonstrates the researcher's understanding (Cronin, Ryan, & Coughlan, 2008).

Chapter 5: Project Methodology

In this chapter different techniques, methods, and features used in this experiment are explored. There are five NLP techniques used in this project: Text Extraction, Text Pre-processing, Text Summarizing, Summarized Text, and Evolution.

Text/Data Extraction: Retrieved the relevant 15 papers from the same business area/research topic from different journals and online portals along with their abstracts and the title has been collected. The data is from an online Journal i.e., Arxiv, Pergamon, ASP. Sample papers have been selected from the area of Customer segmentation.

Text Pre-processing: It is the most vital part of any analysis. Considering a few important pre-processing steps, the below-mentioned techniques have been used.

1. **Stopwords Removal** - Stopwords are meaningless and repeated words that do not contribute to the semantics of the statement. It should be removed.

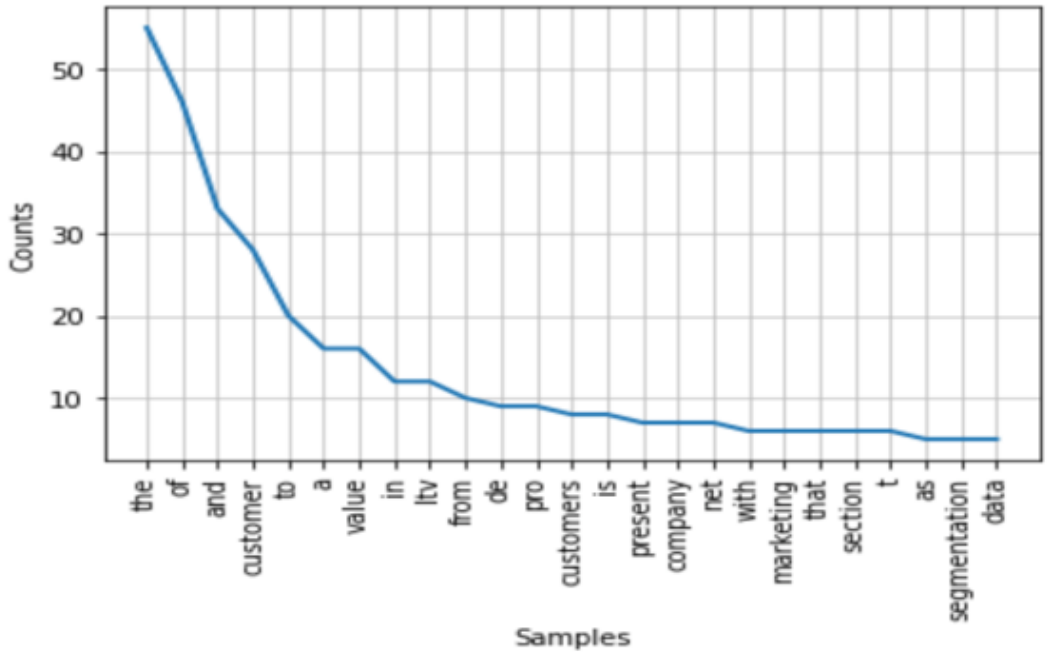


Figure No. 5.1: Word Frequency

In Figure No. 5.1, Word frequency provides the number of times a word occurs.

2. **Symbol Removal** – Reviews generally contain symbols like @,#,$ with no contribution towards analyzing the sentiments. So, it should be removed.
3. **Contractions and Annotation Removal** – Contractions and annotations like ‘shouldn’t’ should be removed with ‘should not.

4. **Exploration** – It is to check the word frequency of the corpus. It gives an idea of what the document is about.

Figure No. 5.2: Word Cloud

In Figure No. 5.2, The volume of each word in a word cloud, a data visualisation technique used to depict text data, shows its frequency or relevance.

Text Summarizing: NLP technique to get the Title feature, Sentence Length, Sentence position, and Sentence Frequency.

Summarized Text: In summarized text researcher could get the Summary of the text based on some predefined criteria.

Evolution: Matrices to evaluate the score of summarized text i.e., Rough, Bert score.

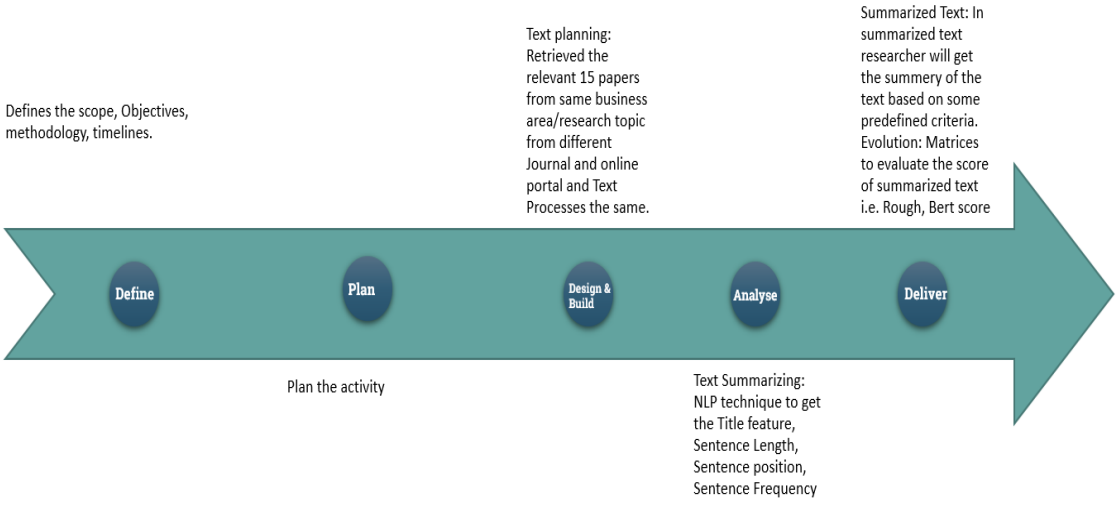


Figure No. 5.3: Project Methodology

In Figure No. 5.3, Here defining the project methodology steps involved in completing the project.

Chapter 6: Business Understanding

The background of the business is described here. The business chosen is an educational institution and students that deal in paper writing. This paper could be helpful for researchers and scholars who are writing papers and spending a lot of time writing literature reviews. The writer usually takes an average of 10 to 15 days to write.

A literature review is the process of evaluating and compiling academic writing on a subject. In a research paper, researchers are gathering what has already been written on that topic; synthesizing the arguments, perspectives, or themes; and summarizing the sources and how they could be applied to the research question.

Problems like the absence of a good literature review, using many unreliable resources, and not including keywords are the weak points of any research paper. The analysis of the research should be done before creating it. If a reader finds the literature review inquisition well-informed and erudite, then only they can go through the research paper.

The following sections make up the article's structure. The literature on business models is examined in the following section in terms of important areas of research, including definition, components, configuration, and typologies; relationships with strategy; business model innovation; business models in emerging markets; theoretical dimensions; and others, including metrics, organisational structure, and leadership. The next parts are devoted to the research's implications and findings.

Chapter 7: Data Understanding

In every study endeavor, data may come from a variety of various sources at different times:

1. Paper submitted on different sources (Arxiv, IEEE).
2. Web Scraping data.
3. Many formats of data (Docx, pdf).

When researchers are prepared to undertake a thorough data analysis, all but the simplest studies must build a system for logging and tracking information. For managing incoming data, different researchers employ various techniques.

Most of the time, consumers want to build a database that enables them to view at any time what data is present and what is still lacking. However, prior knowledge of such systems is necessary. Additionally, it is crucial that the data analyst keep the original data records for a decent period of time, including returned surveys, field notes, test protocols, and so on.

Such records are typically maintained by professional researchers for a minimum of 5-7 years. In a data archive, the original data for significant or pricey investigations may be stored. An analysis result should always be traceable by the data analyst to the original forms on which the data was collected. Effective research records preservation requires a database for recording incoming data.

Chapter 8: Data Preparation

Data preparation includes reviewing or logging the data, confirming the data for accuracy, entering the data into a computer, transforming the data, and building and documenting a database structure that combines the various measures.

8.3 Text Extraction

In this process, the researcher has to retrieve the relevant content from two papers and provide the original text for the pre-processing stage. For demonstration, two papers are chosen for text summarization.

```
' \n \nManagement and maintain of customer relationship have always played a vital role to provide \nbusiness inte
lligence to organizations to build, manage and develop valuable long term customer \nrelationships. The importance
of treating customers as an organizations main asset is increasing in \nvalue in present day and era. Organization
s have an interest to invest in the development of \ncustomer acquisition, maintenance and development strategies.
The business intelligence has a \nvital role to play in allowing companies to use technical expertise to gain bett
er customer \nknowledge and Programs for outreach. By using clustering techniques like k-means, customers \nwith s
imilar means are clustered together. Customer segmentation helps the marketing team to \nrecognize and expose diff
erent customer segments that think differently and follow different \npurchasing strategies. Customer segmentation
helps in figuring out the customers who vary in \nterms of preferences, expectations, desires and attributes. The
main purpose of performing \ncustomer segmentation is to group people, who have similar interest so that the marke
ting team \ncan converge in an effective marketing plan. Clustering is an iterative process of knowledge \ndiscove
ry from vast amounts of raw and unorganized data. Clustering is a type of exploratory data \nmining that is used i
n many applications, such as machine learning, classification and pattern \nrecognition. \n \n \n \n \n \n \n \n \
n \n \n \n \n \n \n \n \n \npropose brief marketing strategies after segmenting custo-\nmer base. This paper is or
ganized as follows. Section \nreviews the previous studies related to customer value. This\npart illustrates the l
imitations of existing studies and\nprepares the background reasons of this paper. Section \nproposes a calculatio
n model for measuring customer value\napplicable to a wireless telecommunication company. We\napply real data of a
wireless company to the model in\nSection . In Section , we perform customer segmentation\nwith the result of cust
omer value derived in Section  and\nproposes brief marketing strategies based upon the result of\ncustomer segment
ation. Finally, Section  concludes this\npaper with the remark on the weaknesses of this study and\nfuture researc
h directions.\n. Related works\n.. The definition of LTV\nCustomer value has been studied under the name of LTV,\nC
ustomer Lifetime Value, Customer Equity, and Customer\nProfitability. The previous researches contain several\ndefin
itions of LTV. The differences between the definitions\nare small. Table  shows the definitions of LTV.\nConsidering
the definitions above, we define LTV as the\nsum of the revenues gained from company's customers over\nthe lifetime
of transactions after the deduction of the total\ncost of attracting, selling, and servicing customers, taking\nin
to account the time value of money. The building block of\nLTV over time frame is shown in Fig. .\nThe horizontal
axis denotes the type of relationship over\ntime frame while vertical, type of customer value toward a\ncompany. A
company forms various relationships according\nto the relationship stages—rudiment, beginning, fosterage,\nand exp
iry stage. A customer also gives a company various\nrevenues, costs, and opportunities and potential benefits.\n..
Models of LTV calculation\nThere are a lot of researches on calculating customer\nvalue. The basic concept of thes
e researches, however,\nfocused on Net Present Value (NPV) obtained from\ncustomers over the lifetime of transacti
ons (Bayo´n,\nGutsche, & Bauer, ; Berger & Nasr, ; Gupta &\nLehmann, ; Roberts & Berger, ). Dwyer ()\ntried to cal
culate LTV through modeling the retention and\nmigration behavior of customers. Focused on making\ndecision of mar
keting invest, Hansotia and Rukstales\n() suggested incremental value modeling using tree\nand regression based ap
proach. Hoekstra and Huizingh\n() also suggested a conceptual LTV model and\ncategorized input data of the model i
nto two types, source\nof interaction data and time frame. Most LTV models stem\nfrom the basic equation, although
we have many other LTV\ncalculation models having various realistic problems. The\nbasic model form based upon the
proposed definition is as\nFig. . The scope of CRM.\nTable \nDefinitions of LTV\nDefinition\nArticle\nThe present val
ue of all future profits\ngenerated from a customer\nGupta and Lehmann ()\nThe net profit or loss to the firm from\na
customer over the entire life of\ntransactions of that customer with\nthe firm\nBerger and Nasr ()\nExpected profits
from customers,\nexclusive of costs related to customer\nmanagement\nBlattberg and Deighton ()\nThe total discount
ed net profit that a\ncustomer generates during her life on\nthe house list\nBitran and Mondschein ()\nThe net pres
ent value of the stream of\ncontributions to profit that result from\ncustomer transactions and contacts\nwith the
company\nPearson ()\nThe net present value of a future stream\nof contributions to overheads and profit\nexpected f
rom the customer\nJackson ()\nThe net present value of all future\ncontributions to overhead and profit\nRoberts an
d Berger ()\nThe net present value of all future\ncontributions to profit and overhead\nexpected from the customer\
nCourtheoux ()\nH. Hwang et al. / Expert Systems with Applications  () —\n\n'
```

Figure No. 8.1: Original Text

In Figure No. 8.1, the original text has been extracted from the papers.

8.4 Text Processing

In this process, the researcher has to clean the text from the original text i.e., remove the URL, special character, number, and provide the clean text for summarizing.

```
'Management and maintain of customer relationship have always played a vital role to provide business intelligence
to organizations to build manage and develop valuable long term customer relationships. The importance of treating
customers as an organizations main asset is increasing in value in present day and era. Organizations have an inte
rest to invest in the development of customer acquisition maintenance and development strategies. The business int
elligence has a vital role to play in allowing companies to use technical expertise to gain better customer knowle
dge and Programs for outreach. By using clustering techniques like k means customers with similar means are cluste
red together. Customer segmentation helps the marketing team to recognize and expose different customer segments t
hat think differently and follow different purchasing strategies. Customer segmentation helps in figuring out the
customers who vary in terms of preferences expectations desires and attributes. The main purpose of performing cus
tomer segmentation is to group people who have similar interest so that the marketing team can converge in an effe
ctive marketing plan. Clustering is an iterative process of knowledge discovery from vast amounts of raw and unorg
anized data. Clustering is a type of exploratory data mining that is used in many applications such as machine lea
rning classification and pattern recognition. propose brief marketing strategies after segmenting custo mer base.
This paper is organized as follows. Section reviews the previous studies related to customer value. This part illu
strates the limitations of existing studies and prepares the background reasons of this paper. Section proposes a
calculation model for measuring customer value applicable to a wireless telecommunication company. We apply real d
ata of a wireless company to the model in Section . In Section we perform customer segmentation with the result of
customer value derived in Section and proposes brief marketing strategies based upon the result of customer segmen
tation. Finally Section concludes this paper with the remark on the weaknesses of this study and future research d
irections. . Related works .. The de nition of LTV Customer value has been studied under the name of LTV Customer
Lifetime Value Customer Equity and Customer Pro tability. The previous researches contain several de nitions of LT
V. The differences between the de nitions are small. Table shows the de nitions of LTV. Considering the de nitions
above we de ne LTV as the sum of the revenues gained from company s customers over the lifetime of transactions af
ter the deduction of the total cost of attracting selling and servicing customers taking into account the time val
ue of money. The building block of LTV over time frame is shown in Fig. . The horizontal axis denotes the type of
relationship over time frame while vertical type of customer value toward a company. A company forms various relat
ionships according to the relationship stages rudiment beginning fosterage and expiry stage. A customer also gives
a company various revenues costs and opportunities and potential bene ts. .. Models of LTV calculation There are a
lot of researches on calculating customer value. The basic concept of these researches however focused on Net Pres
ent Value NPV obtained from customers over the lifetime of transactions Bayo n Gutsche Bauer Berger Nasr Gupta Leh
mann Roberts Berger . Dwyer tried to calculate LTV through modeling the retention and migration behavior of custom
ers. Focused on making decision of marketing invest Hansotia and Rukstales suggested incremental value modeling us
ing tree and regression based approach. Hoekstra and Huizingh also suggested a conceptual LTV model and categorize
d input data of the model into two types source of interaction data and time frame. Most LTV models stem from the
basic equation although we have many other LTV calculation models having various realistic problems. The basic mod
el form based upon the proposed de nition is as Fig. . The scope of CRM. Table De nitions of LTV De nition Article
The present value of all future pro ts generated from a customer Gupta and Lehmann The net pro t or loss to the rm
from a customer over the entire life of transactions of that customer with the rm Berger and Nasr Expected pro ts
from customers exclusive of costs related to customer management Blattberg and Deighton The total discounted net p
ro t that a customer generates during her life on the house list Bitran and Mondschein The net present value of th
e stream of contributions to pro t that result from customer transactions and contacts with the company Pearson Th
e net present value of a future stream of contributions to overheads and pro t expected from the customer Jackson
The net present value of all future contributions to overhead and pro t Roberts and Berger The net present value o
f all future contributions to pro t and overhead expected from the customer Courtheoux H. Hwang et al. Expert Syst
ems with Applications '
```

Figure No. 8.2: Process Text Date

In Figure No. 8.2, the original text has been cleaned using regular expressions.

In the data preparation data has been taken from different papers, with the help of different text cleaning methods like stop words removal, punctuation marks, numbers, URL, special characters, and other non-English words were removed to make the data clean for further processing. Regular expression has also been used to check different frequently occurring patterns in the text to replace or remove them. This clean text can be used with different modeling techniques for text summarization.

Chapter 9: Modelling

Data modeling is the way data is being evaluated. An effective evaluation requires preciseness. Therefore, seven NLP techniques have been taken into account. These include the procedures of Text Rank, Lex Rank, LSA, Luhn, KL Sum, BERT, GPT-2 .

The aim is to understand the underlying efficiency of all these and any problems which occur that could result in the negation of the same. This would lead to a narrower process of analysis. Below are the results of Text Rank, Lex Rank, LSA, Luhn, KL Sum, BERT, and GPT-2.

Text Rank:

```
'Section reviews the previous studies related to customer value. Section proposes a calculation model for measurin
g customer value applicable to a wireless telecommunication company. In Section we perform customer segmentation w
ith the result of customer value derived in Section and proposes brief marketing strategies based upon the result
of customer segmentation. Considering the de nitions above we de ne LTV as the sum of the revenues gained from com
pany s customers over the lifetime of transactions after the deduction of the total cost of attracting selling and
servicing customers taking into account the time value of money. .. Models of LTV calculation There are a lot of r
esearches on calculating customer value. The basic concept of these researches however focused on Net Present Valu
e NPV obtained from customers over the lifetime of transactions Bayo n Gutsche Bauer Berger Nasr Gupta Lehmann Rob
erts Berger . The basic model form based upon the proposed de nition is as Fig.'
```

Figure No. 9.1: Summary using Text Rank

In Figure No. 9.1, getting a summary of the original text using the Text Rank algorithm.

Latent Semantic Analysis:

```
'Customer segmentation helps the marketing team to recognize and expose different customer segments that think dif
ferently and follow different purchasing strategies.The main purpose of performing customer segmentation is to gro
up people who have similar interest so that the marketing team can converge in an effective marketing plan.propose
brief marketing strategies after segmenting custo mer base.Section reviews the previous studies related to custome
r value.We apply real data of a wireless company to the model in Section .The previous researches contain several
de nitions of LTV.The building block of LTV over time frame is shown in Fig.A company forms various relationships
according to the relationship stages rudiment beginning fosterage and expiry stage.The basic concept of these rese
arches however focused on Net Present Value NPV obtained from customers over the lifetime of transactions Bayo n G
utsche Bauer Berger Nasr Gupta Lehmann Roberts Berger .Dwyer tried to calculate LTV through modeling the retention
and migration behavior of customers.'
```

Figure No. 9.2: Summary using LSA

In Figure No. 9.2, getting a summary of the original text using the LSA algorithm.

Lex Rank:

```
'Management and maintain of customer relationship have always played a vital role to provide business intelligence
to organizations to build manage and develop valuable long term customer relationships.Clustering is a type of exp
loratory data mining that is used in many applications such as machine learning classification and pattern recogni
tion.This paper is organized as follows.Section reviews the previous studies related to customer value.Section pro
poses a calculation model for measuring customer value applicable to a wireless telecommunication company.In Secti
on we perform customer segmentation with the result of customer value derived in Section and proposes brief market
ing strategies based upon the result of customer segmentation.The de nition of LTV Customer value has been studied
under the name of LTV Customer Lifetime Value Customer Equity and Customer Pro tability.Considering the de nitions
above we de ne LTV as the sum of the revenues gained from company s customers over the lifetime of transactions af
ter the deduction of the total cost of attracting selling and servicing customers taking into account the time val
ue of money... Models of LTV calculation There are a lot of researches on calculating customer value.Table De niti
ons of LTV De nition Article The present value of all future pro ts generated from a customer Gupta and Lehmann Th
e net pro t or loss to the rm from a customer over the entire life of transactions of that customer with the rm Be
rger and Nasr Expected pro ts from customers exclusive of costs related to customer management Blattberg and Deigh
ton The total discounted net pro t that a customer generates during her life on the house list Bitran and Mondsche
in The net present value of the stream of contributions to pro t that result from customer transactions and contac
ts with the company Pearson The net present value of a future stream of contributions to overheads and pro t expec
ted from the customer Jackson The net present value of all future contributions to overhead and pro t Roberts and
Berger The net present value of all future contributions to pro t and overhead expected from the customer Courtheo
ux H. Hwang et al.'
```

Figure No. 9.3: Summary using Lex Rank

In Figure No. 9.3, getting a summary of the original text using the Lex Rank algorithm.

Luhn:

```
'Management and maintain of customer relationship have always played a vital role to provide business intelligence
to organizations to build manage and develop valuable long term customer relationships.Organizations have an inter
est to invest in the development of customer acquisition maintenance and development strategies.The main purpose o
f performing customer segmentation is to group people who have similar interest so that the marketing team can con
verge in an effective marketing plan.In Section we perform customer segmentation with the result of customer value
derived in Section and proposes brief marketing strategies based upon the result of customer segmentation.The de n
ition of LTV Customer value has been studied under the name of LTV Customer Lifetime Value Customer Equity and Cus
tomer Pro tability.Considering the de nitions above we de ne LTV as the sum of the revenues gained from company s
customers over the lifetime of transactions after the deduction of the total cost of attracting selling and servic
ing customers taking into account the time value of money.The horizontal axis denotes the type of relationship ove
r time frame while vertical type of customer value toward a company.The basic concept of these researches however
focused on Net Present Value NPV obtained from customers over the lifetime of transactions Bayo n Gutsche Bauer Be
rger Nasr Gupta Lehmann Roberts Berger .Hoekstra and Huizingh also suggested a conceptual LTV model and categorize
d input data of the model into two types source of interaction data and time frame.Table De nitions of LTV De niti
on Article The present value of all future pro ts generated from a customer Gupta and Lehmann The net pro t or los
s to the rm from a customer over the entire life of transactions of that customer with the rm Berger and Nasr Expe
cted pro ts from customers exclusive of costs related to customer management Blattberg and Deighton The total disc
ounted net pro t that a customer generates during her life on the house list Bitran and Mondschein The net present
value of the stream of contributions to pro t that result from customer transactions and contacts with the company
Pearson The net present value of a future stream of contributions to overheads and pro t expected from the custome
r Jackson The net present value of all future contributions to overhead and pro t Roberts and Berger The net prese
nt value of all future contributions to pro t and overhead expected from the customer Courtheoux H. Hwang et al.'
```

Figure No. 9.4: Summary using Luhn

In Figure No. 9.4, getting a summary of the original text using the Luhn algorithm.

KL-Sum:

```
'This part illustrates the limitations of existing studies and prepares the background reasons of this paper.Final
ly Section concludes this paper with the remark on the weaknesses of this study and future research directions..Re
lated works ...The horizontal axis denotes the type of relationship over time frame while vertical type of custome
r value toward a company.A company forms various relationships according to the relationship stages rudiment begin
ning fosterage and expiry stage.A customer also gives a company various revenues costs and opportunities and poten
tial bene ts.The scope of CRM.Table De nitions of LTV De nition Article The present value of all future pro ts gen
erated from a customer Gupta and Lehmann The net pro t or loss to the rm from a customer over the entire life of t
ransactions of that customer with the rm Berger and Nasr Expected pro ts from customers exclusive of costs related
to customer management Blattberg and Deighton The total discounted net pro t that a customer generates during her
life on the house list Bitran and Mondschein The net present value of the stream of contributions to pro t that re
sult from customer transactions and contacts with the company Pearson The net present value of a future stream of
contributions to overheads and pro t expected from the customer Jackson The net present value of all future contri
butions to overhead and pro t Roberts and Berger The net present value of all future contributions to pro t and ov
erhead expected from the customer Courtheoux H. Hwang et al.'
```

Figure No. 9.5: Summary using KL-Sum

In Figure No. 9.5, getting a summary of the original text using the KL-Sum algorithm.

BERT:

```
'Management and maintain of customer relationship have always played a vital role to provide business intelligence
to organizations to build manage and develop valuable long term customer relationships. The importance of treating
customers as an organizations main asset is increasing in value in present day and era. We apply real data of a wi
reless company to the model in Section . In Section we perform customer segmentation with the result of customer v
alue derived in Section and proposes brief marketing strategies based upon the result of customer segmentation. Re
lated works .. The de nition of LTV Customer value has been studied under the name of LTV Customer Lifetime Value
Customer Equity and Customer Pro tability. The previous researches contain several de nitions of LTV. The horizont
al axis denotes the type of relationship over time frame while vertical type of customer value toward a company.'
```

Figure No. 9.6: Summary using Bert

In Figure No. 9.6, getting a summary of the original text using the BERT algorithm.

GPT-2:

```
'Management and maintain of customer relationship have always played a vital role to provide business intelligence
to organizations to build manage and develop valuable long term customer relationships. The main purpose of perfor
ming customer segmentation is to group people who have similar interest so that the marketing team can converge in
an effective marketing plan. Clustering is an iterative process of knowledge discovery from vast amounts of raw an
d unorganized data. The previous researches contain several de nitions of LTV. The basic concept of these research
es however focused on Net Present Value NPV obtained from customers over the lifetime of transactions Bayo n Gutsc
he Bauer Berger Nasr Gupta Lehmann Roberts Berger . Dwyer tried to calculate LTV through modeling the retention an
d migration behavior of customers. The basic model form based upon the proposed de nition is as Fig. .'
```

Figure No. 9.7: Summary using GPT-2

In Figure No. 9.7, getting a summary of the original text using the GPT-2 algorithm.

Chapter 10: Model Evaluation

The evaluation of the same has been performed keeping in mind the pros and cons of every procedure. The proposed model has been run across the different models for evaluation purposes.

There are seven text analytics algorithms (Text Rank, Lex Rank, LSA, Luhn, KL Sum, BERT, GPT-2) and three evaluation matrices (Rough-1, Bert Score) that have been used. Of these, the most suitable has resulted in optimal results.

Evaluation research proposals are used to negotiate a study commission, project contract, and/or to carry out the actual evaluation research study. Once a study is finished, a written assessment report must be submitted.

Text Rank Model:

- **Bert Score:**

	Score
Bert	0.22

Table No. 10.1: Bert Score for Text Rank

In Table No. 10.1, getting Bert score for the Text Rank algorithm.

- **Rouge-L:**

	ROUGE-1	ROUGE-2	ROUGE-L
r	0.27	0.2	0.27
p	1	0.97	1
f	0.42	0.33	0.42

Table No. 10.2: Rough Score for Text Rank

In Table No. 10.2, getting the Rouge score for the Text Rank algorithm.

Latent Semantic Analysis Model:

- **Bert Score:**

	Score
Bert	0.23

Table No. 10.3: Bert Score for LSA

In Table No. 10.3, getting Bert score for the LSA algorithm.

- **Rouge-L:**

	ROUGE-1	ROUGE-2	ROUGE-L
r	0.33	0.22	0.33
p	1	0.96	1
f	0.49	0.36	0.49

Table No. 10.4: Rouge Score for LSA

In Table No. 10.4, getting the Rouge score for the LSA algorithm.

Lex Rank Model:

- **Bert Score:**

	Score
Bert	0.24

Table No. 10.5: Bert Score for Lex Rank

In Table No. 10.5, getting Bert score for the Lex Rank algorithm.

- **Rouge-L**

	ROUGE-1	ROUGE-2	ROUGE-L
r	0.48	0.42	0.47
p	1	0.97	1
f	0.64	0.59	0.65

Table No. 10.6: Rouge Score for Lex Rank

In Table No. 10.6, getting the Rouge score for the Lex Rank algorithm.

Luhn Model:

- **Bert Score:**

	Score
Bert	0.24

Table No. 10.7: Bert Score for Luhn

In Table No. 10.7, getting Bert score for the Luhn algorithm.

- **Rouge-L:**

	ROUGE-1	ROUGE-2	ROUGE-L
r	0.54	0.5	0.54
p	1	0.98	1
f	0.7	0.66	0.7

Table No. 10.8: Rouge Score for Luhn

In Table No. 10.8, getting the Rouge score for the Luhn algorithm.

KL-Sum Model:

- **Bert Score:**

	Score
Bert	-0.03

Table No. 10.9: Bert Score for KL-Sum

In Table No. 10.9, getting Bert score for the KL-Sum algorithm.

- **Rouge-L:**

	ROUGE-1	ROUGE-2	ROUGE-L
r	0.34	0.3	0.34
p	1	0.99	1
f	0.51	0.46	0.51

Table No. 10.10: Rouge Score for KL-Sum

In Table No. 10.10, getting the Rouge score for the KL-Sum algorithm.

BERT Model:

- **Bert Score:**

	Score
Bert	0.42

Table No. 10.11: Bert Score for Bert

In Table No. 10.11, getting the Bert score for the Bert algorithm.

- **Rouge-L:**

	ROUGE-1	ROUGE-2	ROUGE-L
r	0.22	0.14	0.22
p	1	0.96	1
f	0.36	0.25	0.36

Table No. 10.12: Rouge Score for Bert

In Table No. 10.12, getting the Rouge score for the Bert algorithm.

GPT-2 Model:

- **Bert Score:**

	Score
Bert	0.28

Table No. 10.13: Bert Score for GPT-2

In Table No. 10.13, getting Bert score for GPT-2 algorithm.

- **Rouge-L:**

	ROUGE-1	ROUGE-2	ROUGE-L
r	0.27	0.17	0.27
p	1	0.98	1
f	0.42	0.29	0.42

Table No. 10.14: Rouge Score for GPT-2

In Table No. 10.14, getting the Rouge score for GPT-2 algorithm.

Chapter 11: Deployment

The model deployment has been done using GPT2 and BERT which summarizes the literature review to give a new literature review.

Users can open the models with the help of the given URL link and upload all the research papers for which users want a text summarization. On submitting the paper user can see the summarized text on the page.

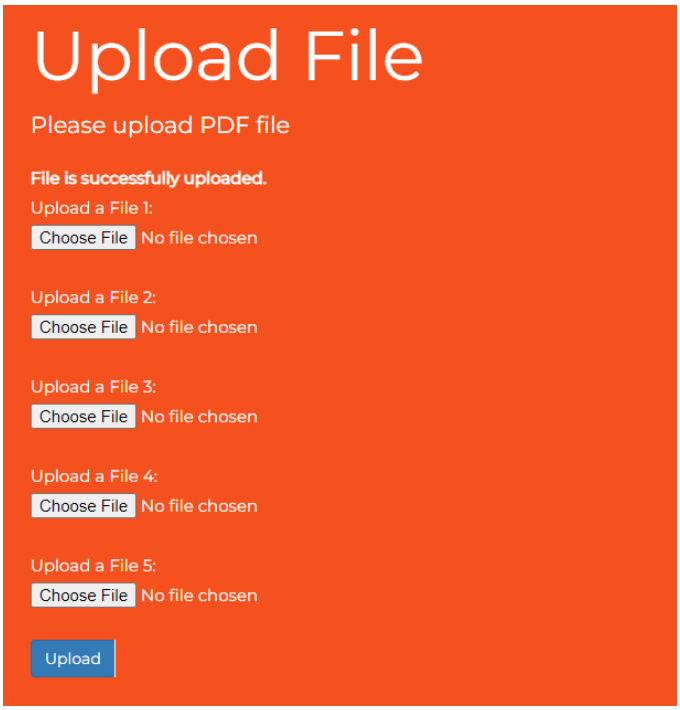


Figure No. 11.1: Webpage for uploading the document

In Figure No. 11.1, Web page for uploading only pdf document.

Chapter 12: Analysis and Result

A researcher can be prepared to evaluate each article (dissect it, find the relevant information inside it), then synthesise the collection of articles once they have collected the articles for a research paper (integrate them and identify the conclusions that can be drawn from the articles as a group). It is advisable to start by reading through the assembled articles to get a sense of what they are about. Most published papers provide an overview or abstract at the start of the document, which could help with this process and enable a researcher to choose whether it is worthwhile to read the entire document or to include it. It could also be helpful at this point to classify and organise the articles initially according to the type of source (Cronin, Ryan, & Coughlan, 2008).

In this research report total of seven different NLP techniques (Text Rank, Lex Rank, LSA, Luhn, KL Sum, BERT, GPT-2) have been used based on their BERT score, transformer-based techniques outperform others. So, there are two techniques GPT2 and BERT which were deployed in the production for the end user uses.

This part of the research enumerates the result and its analysis: of the repercussions of the achieved result. This is inclusive of the outcome of the proposed methodology and covers descriptive work.
The outcome was best when the GPT-2 was adopted as the procedure as it showed a better performance in the field of text generation in the field of **Automated Literature Review with Transformers.**

Bert text summarization report:
The backbone of BERT work is based on transformer, which is utilised to get beyond the long-term dependencies that limit RNN and other neural networks. It is a naturally bidirectional pre-trained model. This pre-trained model can be simply adjusted to carry out the given NLP tasks, in our example, Summarization.

'Management and maintain of customer relationship have always played a vital role to provide business intelligence to organizations to build manage and develop valuable long term customer relationships. The importance of treating customers as an organizations main asset is increasing in value in present day and era. We apply real data of a wi reless company to the model in Section . In Section we perform customer segmentation with the result of customer v alue derived in Section and proposes brief marketing strategies based upon the result of customer segmentation. Re lated works .. The de nition of LTV Customer value has been studied under the name of LTV Customer Lifetime Value Customer Equity and Customer Pro tability. The previous researches contain several de nitions of LTV. The horizont al axis denotes the type of relationship over time frame while vertical type of customer value toward a company.'

Figure No. 12.1: Summary for Bert

In Figure No. 12.1, getting a summary of the original text using the BERT algorithm.

GPT-2 text generation report: GPT-2 is a seq2seq model, it can also be fine-tuned for the task of text summarization. Here the format of data is very similar to the translation task- "*text = summary".*

'Management and maintain of customer relationship have always played a vital role to provide business intelligence to organizations to build manage and develop valuable long term customer relationships. The main purpose of perfor ming customer segmentation is to group people who have similar interest so that the marketing team can converge in an effective marketing plan. Clustering is an iterative process of knowledge discovery from vast amounts of raw an d unorganized data. The previous researches contain several de nitions of LTV. The basic concept of these research es however focused on Net Present Value NPV obtained from customers over the lifetime of transactions Bayo n Gutsc he Bauer Berger Nasr Gupta Lehmann Roberts Berger . Dwyer tried to calculate LTV through modeling the retention an d migration behavior of customers. The basic model form based upon the proposed de nition is as Fig. .'

Figure No. 12.2: Summary for GPT-2

In Figure No. 12.2, getting a summary of the original text using the GPT-2 algorithm.

Validation: Two different research papers on the topic of Customer segmentation have been taken for the text summarization:

1. Customer segmentation using clustering algorithms.
2. Customer Segmentation Technique on Ecommerce

After the text summarization researcher could get the context of it in a small summary para with the citation of each included in it, so he can refer to the same in their research. As of now in this work researcher can get the name of the research papers in place of the Author's name and year as per American Psychological Association 7th Ed. Citation style. Later in the next phases, other citation styles could also be included.

'Literature Review Customer Segmentation Over the years as there is very strong competition in the business world the organizations have to enhance their profits and business by satisfying the demands of their customers and attract new customers according to their needs. The identification of customers and satisfying the demands of each customer is a very complex task. According to customer segmentation is a strategy of dividing the market into homogenous groups.': ['data/**Customer segmentation using clustering algorithms**.pdf]'

'Review on Customer Segmentation Technique on Ecommerce Ecommerce transactions are no longer a new thing. In marketing personalization technique can be used to get potential customers in a case to boost sales. Duration when seeing the product can be used as customer interest in the product so that it can be used as a variable in customer segmentation. Keywords Ecommerce Customer Segmentation Personalization. It will generate more profits for the company. Several researchers discuss the customer segmentation method on their papers such as Magento who used several variables to perform customer segmentation namely transaction variable product variable geographic variable hobbies variable and page viewed variable Baer and Colica discuss customer segmentation methods of Business Rule Quantile membership Supervised Clustering Unsupervised Clustering Customer Profiling RFM Cell Classification Grouping Customer Likeness Clustering and Purchase Affinity Clustering.': ['data/**Review on Customer Segmentation Technique on Ecommerce.pdf**']

Figure No. 12.3: Summary using Bert against each paper

In Figure No. 12.3, summarised the literature review of all the papers using BERT with paper name.

Chapter 13: Conclusions and Future Scope

The literature review of research papers has been summarized to create a new literature review. Here normal research paper (Not IEEE) has been taken under consideration to develop the model.

Future work can be considered IEEE format research work and high-end models like GPT3 models could be used.

Bibliography

Agrawal, S., Foster, G., Freitag, M., & Cherry, C. (2021). Assessing Reference-Free Peer Evaluation for Machine Translation. *arxiv*, 1-2.

Barrios, F., L´opez, F., Argerich, L., & Wachenchauzer, R. (2016). Variations of the Similarity Function of TextRank for Automated Summarization. *arXiv*, 1-3.

Carnwell, R., & Daly, W. (2001). Strategies for the Construction of a Critical Review of the Literature. *Nurse Education in Practice*, 1: 57-63.

Celikyilmaz, A., Clark, E., & Gao, J. (2021). Evaluation of Text Generation: A Survey. *arxiv*, 14.

Cooper, H. M. (1988). Organizing Knowledge Syntheses. *Knowledge in Society*, 1: 104-126.

Cronin, P., Ryan, F., & Coughlan, M. (2008). Undertaking a literature review: a step-by-step approach. *British Journal of Nursing*, 17(1): 38-43.

Ganesan, K. (2018). ROUGE 2.0: Updated and Improved Measures for Evaluation of Summarization Tasks. *arxiv*, 1-3.

Hart, C. (1998). Fault Prediction Performance in Software Engineering. *London: Sage Publications*.

Kennedy, & M, M. (2017). Defining a Literature. *Educational Researcher*, 139-147.

Khandelwal, U., Clark, K., Jurafsky, D., & Kaiser, Ł. (2019). Sample Efficient Text Summarization Using a Single Pre-Trained. *arxiv*, 1-2.

Liu, Y. (2019). Fine-tune BERT for Extractive Summarization. *arxiv*, 1-5.

Mallick, C., Das, A. K., Dutta, M., Das, A. K., & Sarkar, A. (2019). Graph-Based Text Summarization Using Modified TextRank. *researchgate*, 1-4.

Miller, D., & Atlanta, G. (2019). Leveraging BERT for Extractive Text Summarization on Lectures. *arxiv*, 1-6.

Okoli, C., & Schabram, K. (2010). A Guide to Conducting a Systematic Literature Review of Information Systems Research. *Sprouts: Working Papers on Information Systems*, 10(26).

Ozsoy, M. G., Alpaslan, F. N., & Cicekli, I. (2011). Text summarization using Latent Semantic Analysis. *SAGE*, 1-3.

Radev, G. E., & R., D. (2004). LexRank:Graph-based Lexical Centrality as Salience in Text Summarization. *jair*, 1-5.

Sally. (2013). A Synthesis Matrix as a Tool for Analysing and Synthesizing Prior Research.

ShafieiBavani, E., Ebrahimi, M., Wong, R., & Chen, F. (2017). A Semantically Motivated Approach to Compute ROUGE Scores. *arxiv*, 1-3.

Tan, B., & Kieuvongngam, V. (2020). Automatic Text Summarization of COVID-19 Medical Research Articles using BERT and GPT-2. *arxiv*, 1-6.

Timmins, F., & McCabe, C. (2005). How to conduct an effective literature review. *Nurs Stand*, 20(11): 41–7.

Uçkan, T., & Karcı, A. (2020). Extractive multi-document text summarization based on graph. *Egyptian Informatics*, 1-6.

Younger, P. (2004). Using the internet to conduct a literature search. *Nurs Stand*, 19 (6): 45-51.

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., & Artzi, Y. (2020). BERTSCORE: EVALUATING TEXT GENERATION WITH BERT. *arxiv*, 1-3.

Appendix

Plagiarism Report[1]

Automated Literature Review with Transformers

by Sanjeev Kumar Jha

Submission date: 26-Aug-2022 12:34PM (UTC+0530)
Submission ID: 1887332468
File name: Automated_Literature_Review_with_Transformers_-_Sanjeev_Jha.docx (3.42M)
Word count: 5787
Character count: 30921

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

Automated Literature Review with Transformers

ORIGINALITY REPORT

13%	8%	1%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.researchgate.net Internet Source	3%
2	Submitted to Cardiff University Student Paper	1%
3	medium.com Internet Source	1%
4	Submitted to University of South Africa Student Paper	1%
5	www.machinelearningplus.com Internet Source	1%
6	Submitted to Institute of Technology, Nirma University Student Paper	1%
7	Submitted to Sunway Education Group Student Paper	1%
8	Submitted to National Institute of Construction Management and Research Student Paper	1%

9	Submitted to Asia Pacific University College of Technology and Innovation (UCTI) Student Paper	1 %
10	Submitted to Liverpool John Moores University Student Paper	<1 %
11	Submitted to University of Zululand Student Paper	<1 %
12	guides.lib.ua.edu Internet Source	<1 %
13	Submitted to Flinders University Student Paper	<1 %
14	en.wikipedia.org Internet Source	<1 %
15	issuu.com Internet Source	<1 %
16	Submitted to Universiti Teknologi MARA Student Paper	<1 %
17	Sepp Hochreiter, Jürgen Schmidhuber. "Long Short-Term Memory", Neural Computation, 1997 Publication	<1 %
18	Submitted to University of College Cork Student Paper	<1 %
	libguides.usc.edu	

19	Internet Source	<1%
20	Submitted to Cyryx College, Maldives Student Paper	<1%
21	Submitted to Global Banking Training Student Paper	<1%
22	Submitted to The University of the West of Scotland Student Paper	<1%

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		

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

Paper Submitted:
Sanjeev Kumar Jha, Ashish Chandra Jha, Dr. J. B. Simha, "Automated Literature Review with Transformers" 9th International Conference, IIM Bangalore **2022.**

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

Automated Literature Review with Transformers

1st Sanjeev Kumar Jha
REVA Academy for Corporate Excellence – RACE, *REVA University*
Bengaluru, India
sanjeev.ba05@reva.edu.in

2nd Ashish Chandra Jha
REVA Academy for Corporate Excellence – RACE, *REVA University*
Bengaluru, India
ashish.ba05@reva.edu.in

3rd Dr. J. B. Simha
REVA Academy for Corporate Excellence – RACE, REVA University
Bengaluru, India
jb.simha@reva.edu.in

***Abstract*— Business research is producing new knowledge at an astounding rate, yet it is still fragmented and heterogeneous. This makes it challenging to stay up to date, be at the forefront of research, and interpret the data in context in a particular area of business research. Because of this, it is more crucial than ever as a research strategy. Traditional literature reviews usually lack thoroughness, rigor and are conducted haphazardly rather than in accordance with a predetermined methodology. As a result, questions about the validity and reliability of these kinds of evaluations may be brought up. Despite the value of conducting Systematic Literature Reviews (SLR) of the literature to pinpoint research needs across various disciplines of study, manually doing SLR is a challenging, multi-stage, and time-consuming process.**

The primary goal of this research topic is to extract the reference paper and summarize it based on existing research papers in the same area and generate a Summary of the text which prevents duplication.

The text has been taken from different papers and after that with the help of different regular expressions collected data was cleaned from unnecessary words or punctuation marks. On the cleaned text different Natural Language Processing techniques (Text Rank, Lex Rank, LSA, Luhn, KL Sum, BERT, GPT-2) have been utilized for the summarization of the text. Out of all NLP techniques it has beenobserved that GPT and BERT were giving the best result.

On a selected topic, different research papers have been explored. Literature reviews of different research papers have been taken and summarized for reference to new researchers in that field. In this way, this model reduces the research time of different researchers and gives them an idea of previous research which has taken place on this topic in recent years.

Keywords: Text Summarization, Transformer, Research Paper, Extractive Summarization, Transformer, NLP

I. Introduction

A literature review looks at information that has been published on a particular subject and, on occasion, information from a certain time period in a particular field. A mature subject addresses the need for a critique of the topic's expanding and more varied body of knowledge as well as prospective reconceptualization. The second kind of it is centered on cutting-edge or unique topics that could profit from a thorough conceptualization and synthesis of the literature. Because of the novelty of these topics and the lack of a comprehensive literature review, A new model or framework, as well as a first or preliminary conception of the subject, are more likely to emerge from the review.There are several potential justifications for doing this, including gathering data for the creation of evidence-based care, both as a step in the research process and as a component of an academic assessment. Conducting it seems to be a challenging task for too many capable students. The most often queries range from what a literature review comprises to how to choose a subject, how many articles to include, and where to start [1].

Literature Review Understanding:
In this sense, it is a scientific publication that presents up-to-date information, including significant findings as well as theoretical and methodological advancements in a certain field. Secondary sources, do not report on new or unusual experimental work, as in literature reviews. These reviews which are most frequently associated with academic literature, are published in scholarly publications, and should not be confused with book reviews, they might be published in the same journal. Evaluations of the literature form the basisfor study in almost every academic field.

It is crucial to think of knowledge as having three layers while discussing a particular subject.

1. The researchers' initial studies, known as primary studies, are carried out and published.
2. Second, assessments of those results often go beyond the original investigations while summarizing and proposing new interpretations based on them.
3. Third, there are unofficial judgments, opinions, and interpretations that get incorporated into the legend around the subject.

It is important to note that this third layer of knowledge is commonly referred to as "true" while writing this, despite just having a tenuous connection to the primary studies and secondary literature evaluations.

The three categories are *evaluative, exploratory, and instrumental*. The systematic review, a fourth form that is sometimes categorised independently, is simply a literature review that is concentrated on a particular research issue and makes an effort to find, evaluate, select, and synthesise all high-quality research data and arguments pertinent to that topic.
A meta-analysis is a systematic review that makes good use of statistical tools to combine data from all chosen studies in order to get a more accurate conclusion. An interactives' objective is to create new knowledge about a subject by analysing, evaluating, and synthesizing the literature in question [2].

*Types of Literature Review***:** The numerous types of literature reviews are listed below:

Argumentative Review
This approach examines the literature only when it is necessary to support or refute an established thesis, a long-held belief, or a philosophical issue. The main objective is to produce a corpus of writing that could serve as a counterargument. Considering how some social science studies are value-laden. For example, immigration regulation and educational reform. Argumentative methods of literary analysis may be appropriate and worthwhile. However, they can cause bias issues when used to generate summary claims like those seen in systematic reviews [2].

Integrative Review
Known as a study style that incorporates, analyses, evaluates, and synthesises representative literature on a subject to create fresh frameworks and viewpoints on the subject. Within the body of literature are all the research that specifically address related or parallel hypotheses. A well-done integrative review satisfies the same standards as primary research in terms of clarity, rigour, and replication [2].

Historical Review
Few things stand apart from historical precedent. Examining studies through time, historical reviews frequently start with the first time an issue, concept, theory, or phenomenon appears in the literature and trace its development within a discipline's scholarship. The objective is to contextualise research within a historical framework to demonstrate knowledge through cutting-edge discoveries and to identify potential future research fields [2].

Methodological Review
The focus of a review is often on the method of analysis rather than the content of what was said. This method offers a framework of comprehension at different levels (theory, substantive fields, research approaches, and data collection and analysis techniques), allowing researchers to draw on a variety of knowledge ranging from the conceptual level to practical documents for use in the fieldwork in many fields of ontological and epistemological consideration as well as for quantitative and qualitative analysis [2].

Systematic Review
In order to find and critically evaluate relevant research, as well as to gather, present, and analyse data from the studies included in the review, this form includes an overview of the currently available evidence pertinent to a clearly defined research question. The main focus is typically on a relatively specific empirical topic, which is frequently presented in a cause-and-effect structure, such as "How much does A contribute to B?" [2].

Theoretical Review
This form's objective is to investigate the body of theory that has been accumulated in regard to a certain subject, idea, theory, or phenomenon in a concrete way. Theoretical analysis aids in identifying current theories, their connections, the depth to which concepts have been researched, and the creation of new hypotheses to be tested. This method is widely used to show that there are no valid theories or that existing theories are insufficient to explain novel or emerging research problems. A single theoretical idea, a whole theory [2].

As there are total of six types of literature reviews i.e., Argumentative, Integrative, Historical, Methodological, systematic, and theoretical. Different research falls into a different category by summarizing it researcher can come to know in which category a particular paper belongs to.

II. Literate Review

The writing of a thesis, dissertation, or journal article, among other graduate and post-graduate student assignments, typically calls for it. A study proposal or prospectus frequently includes a review of the literature (the agreement signed before a researcher officially starts writing a dissertation or thesis). It may simply be a summary of relevant sources. The key points of the source are summarised in a summary, whereas the information is reorganised or rearranged in a synthesis to help with how to analyse a research paper.

This paper creates a list of previously published articles on a particular subject. A full academic work or a portion of one.In any case, the goal of this is to provide the researcher/authorand the audience with a broad overview of the material that is currently accessible on the topic at hand. A thoroughreview of the literature helps ensure that an appropriate theoretical framework and/or research methods were used, aswell as that a valid study issue was addressed. In other words,this gives the reader a perspective by setting the current workwithin the context of the pertinent literature. In these circumstances, the review typically comes before the sections discussing the work's procedures and results.

A review article can be included as a literature review. It is a scientific journal that delivers up-to-date information, together with significant findings and theoretical and methodological contributions to a particular field. Literature reviews are an example of a secondary source that does not present new or original experimental work. Academic publications sometimes contain such reviews, which are frequently associated with academic literature. Practically every academic discipline's research starts with a literature review.

Although some of the earliest reports mentioned in the literature were vocal, written reports make up the vast majority of reports. scholarly work that may be methodological, theoretical, empirical, or critical in nature. Second, it makes an effort to summarise, analyse, evaluate, make clear, and/or include the content of original reports [3]. A complete summary and critical analysis of the existing research and non-research literature on the subject at hand constitutes a literature review [1] [4]. This serves as a foundation for another goal, such as providing evidence for further research into the field, while also keeping the reader informed about recent publications on the subject. A good literature review gathers data from several sources on a

particular subject. It should have a distinct search and selection process [5].

It differs from a report on scholarly research. Establishing a novel argument is one of the main objectives of an academic research paper, and it might be one of its components. In a research paper, the researcher used the literature to provide the groundwork and provide evidence for novel insight. Contrarily, this aims to summarise and synthesise the thoughts and arguments of others without adding anything new.

There are several uses for this, and the vast majority of them are found in a primary research article that provides the theoretical framework for the article's main investigation. It provides the framework for the remaining parts of an academic piece. It clearly communicates to the reader the importance of earlier work while explaining the nature and content of the current body of knowledge [6]. The review must synthesise the available information and provide a scientific critique of the theory in order to contribute to the work as an academic paper. It cannot simply restate the subject matter [6].

III. METHODOLOGY

In this section, will explore the different techniques,methods, and features used in this experiment and also will dividethe section into two sub sections: data exploration and pre-processing and model building.

Text/Data Extraction: Retrieved the relevant 15 papersfrom the same business area/research topic from different journals and online portals along with their abstracts and thetitle has been collected. The data is from an online Journal i.e., Arxiv, Pergamon. Sample papers have been selected from the area of Customer segmentation.

Text Pre-processing: It is the most vital part of any analysis. Considering a few important pre-processing steps, the below-mentioned techniques have been used.

1. *Stopwords Removal* - Stopwords are meaningless and repeated words that do not contribute to the semantics of the statement. It should be removed.

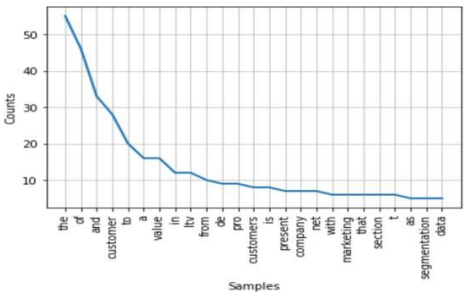


Figure No. 3.1: Word Frequency

2. *Symbol Removal* – Reviews generally contain symbols like @,#,$ with no contribution towards analyzing the sentiments. So, it should be removed.
3. *Contractions and Annotation Removal* – Contractionsand annotations like ‘shouldn’t’ should be removed with ‘should not.
4. *Exploration* – It is to check the word frequency of the corpus. It gives an idea of what the document is about.

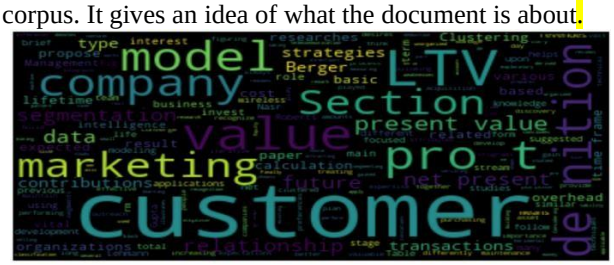


Figure No. 3.2: Word Cloud

Text Summarizing: NLP technique to get the Title feature, Sentence Length, Sentence position, and Sentence Frequency.

Summarized Text: In summarized text researcher could get the Summary of the text based on some predefined criteria.

Seven algorithms have been identified as having the potential to Generate it. These are as follows:

Text Rank: It is a graph-based text processing ranking model that demonstrates how this model may be used effectively in natural language applications. When utilising anunsupervised method to extract phrases and keywords, it is really helpful. An extractive summarization method is TextRank. It is predicated on the idea that words with higherfrequency have greater significance [7] [8].

1. An extractive summarization method is TextRank.
2. It is predicated on the idea that words with higher frequency have greater significance.
3. Therefore, it is crucial to pay attention to sentences with many common words.
4. The system awards grades to each sentence in the text based on this.

Lex Rank: It is a different PageRank method offspring with the text summarising tool TextRank as a sibling. A sentence is likely to be considered important if it can be related to several other sentences in the text. Finding a sentence that is supported by other, comparable sentences and is scored higher is how LexRank works. The higher the rank of the sentence is, the higher the priority of being included within the summarized text [9] [10].

An unsupervised learning approach called latent semantic analysis can be applied to extractive text summarization.By applying Singular Value Decomposition (SVD) to the matrix of term-document frequency, it extracts phrases that are semantically significant.

Latent Semantic Analysis (LSA): it is a technique for examining the connections between a collection of documents and the terms they include by creating a collection of concepts pertaining to both the documents and the terms. By using SVD on the grid of term-document frequency, it extracts phrases that are semantically significant. Latent

Semantic Analysis is a generally applicable unsupervised learning approach for extractive text summarization [11][10].

1. A sentence has a high likelihood of being significant if it is comparable to numerous other sentences in the text.
2. According to LexRank's methodology, a sentence is ranked higher if other sentences that are similar to it recommend it.
3. The priority of being included in the summary text increases with rank.

Luhn: Based on Term Frequency-Inverse Document Frequency (TF-IDF) method for summarising Luhn When both very low-frequency words and highly common words (stopwords), approach is quite helpful. It is used to separate sentence ranks, and the top-ranking sentences are featured in the summary [10].

1. The technique of the Luhn Summarization algorithm is based on TF-IDF (Term Frequency-Inverse Document Frequency). When both highly frequent words (stopwords) and very low frequency words (VLF) are insignificant, it is helpful.
2. Accordingly, sentence scoring is done, and the best-scoring sentences are included in the summary.

KL Sum: It is an extractive strategy that chooses phrases based on how frequently words appear in the original text. It seeks to reduce the KL-divergence threshold (learn more). To reduce the KL-divergence, it employs a greedy optimization strategy and keeps adding sentences [10].

1. The KL-Sum algorithm is another extraction technique.
2. Based on how closely the word distribution resembles the original text, it chooses which sentences to use.
3. It seeks to reduce the KL-divergence threshold (learn more).
4. It employs a greedy optimization strategy and continues to add phrases as long as the KL-divergence remains low.

Bidirectional Encoder Representations from Transformers (BERT): BERT is a transformer that is used to get around the long-term dependencies that Recurrent Neural Networks (RNN) and other neural networks have. It is a pre-trained naturally bidirectional model. This pre- trained model can easily complete the given NLP tasks because it is highly adaptable, which is in our casesummarization [6] [12] [13].

BERT is a transformer that is used to get around the long-term dependencies that RNN and other neural networks have. It is a pre-trained model that is naturally bidirectional. This pre-trained model can be adjusted to easily carry out the specified NLP tasks, Summarization in our case.

Generative Pretrained Transformer 2 (GPT-2): GPT-2 isa seq2seq model that can be altered to carry out the task of text summarization. Here the format of data is very similar tothe translation task- "*text = summary"* [13] [14].

When the researcher searches the literature review in different research papers it's not an easy task to understand the context of it. Usually, it takes a considerable time to read and understand it fully. It is quite difficult to summarize the whole literature review in a few words manually by a research scholar. When they need to refer to a greater number of papers on any topic it becomes a very time-consuming and tough task. So, there was a need for a method or model whichcan resolve this issue and provide some relief to the research scholars.

The text summarization model takes a lot of text for processing that's why it needs a high-end configurations system with high computation intensive CPU\GPU otherwise for the normal system it's a time-consuming task.

*Evolution***:** Matrices to evaluate the score of summarized text i.e., Rough, Bert score.

Recall-Oriented Understudy for Gisting Evaluation-N (Rough-N): Between the model-generated text and references RAUGE-N measures the number of matching n- grams. ROUGE score can be used for scoring [15] [16].

Bert Score: it is an automatic evaluation metric for text generation from existing text. The main advantage of using this metric is to calculate a score of comparable ness between each token in the candidate sentence and each token in the reference sentence [17].

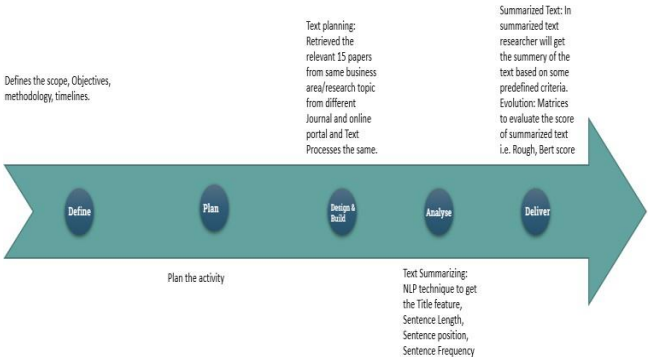


Figure No. 3.4: Methodology

IV. RESULTS AND DISCUSSION

This part of the research enumerates the result and its analysis: of the repercussions of the achieved result. This is inclusive of the outcome of the proposed methodology and covers descriptive work.

The outcome was best when the GPT-2/ GPT-3 was adopted as the procedure as it showed a better performance in the field of text generation in the field of Automated Literature Review with Transformers.

Validation: Two different research papers on the topic of Customer segmentation have been taken for the text summarization:

1. Customer Segmentation using clustering algorithms.
2. Customer Segmentation Technique on Ecommerce

'Literature Review Customer Segmentation Over the years as there is very strong competition in the business world the organizations have to enhance their profits and business by satisfying the demands of their customers and attract new customers according to their needs. The identification of customers and satisfying the demands of each customer is a very complex task. According to customer segmentation is a strategy of dividing the market into homogenous groups.': ['data/**Customer segmentation using clustering algorithms**.pdf]'

'Review on Customer Segmentation Technique on Ecommerce Ecommerce transactions are no longer a new thing. In marketing personalization technique can be used to get potential customers in a case to boost sales. Duration when seeing the product can be used as customer interest in the product so that it can be used as a variable in customer segmentation. Keywords Ecommerce Customer Segmentation Personalization. It will generate more profits for the company. Several researchers discuss the customer segmentation method on their papers such as Magento who used several variables to perform customer segmentation namely transaction variable product variable geographic variable hobbies variable and page viewed variable Baer and Colica discuss customer segmentation methods of Business Rule Quantile membership Supervised Clustering Unsupervised Clustering Customer Profiling RFM Cell Classification Grouping Customer Likeness Clustering and Purchase Affinity Clustering.': ['data/**Review on Customer Segmentation Technique on Ecommerce.pdf**']

Figure No. 4: Summary using Bert against each paper

Score:

Score										
	Rouge									BERT
	Recall			Presission			F-Score			F-Score
	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L	ROUGE-1	ROUGE-2	ROUGE-L	
Model Name										
TextRank	0.27	0.2	0.27	1	0.97	1	0.42	0.33	0.42	0.22
LexRank	0.48	0.42	0.47	1	0.97	1	0.64	0.59	0.65	0.24
LSA	0.33	0.22	0.33	1	0.96	1	0.49	0.36	0.49	0.23
Luhan	0.54	0.5	0.54	1	0.98	1	0.7	0.66	0.7	0.24
KL-sum	0.34	0.3	0.34	1	0.99	1	0.42	0.33	0.42	-0.03
Bert	0.22	0.14	0.22	1	0.96	1	0.36	0.25	0.36	0.42
Gpt-2	0.27	0.17	0.27	1	0.98	1	0.42	0.29	0.42	0.28

After the text summarization researcher could get the context of it in a small summary para with the citation of each included in it, so he can refer to the same in their research.

V. REFERENCES

[1] P. Cronin, F. Ryan and M. Coughlan, "Undertaking a literature review: a step-by-step approach," *British Journal of Nursing*, pp. 17(1): 38-43, 2008.

[2] Kennedy and M. M, "Defining a Literature.," *Educational Researcher,* pp. 139-147, 2017.

[3] H. M. Cooper, "Organizing Knowledge Syntheses," *Knowledge in Society*, pp. 1: 104-126, 1988.

[4] C. Hart, "Fault Prediction Performance in Software Engineering," *London: Sage Publications,* 1998.

[5] R. Carnwell and W. Daly, "Strategies for the Construction of a Critical Review of the Literature.," *Nurse Education in Practice*, pp. 1: 57-63, 2001.

[6] D. Miller and G. Atlanta, "Leveraging BERT for Extractive Text Summarization on Lectures," *arxiv*, pp. 1-6, 2019.

[7] F. Barrios, F. L´opez, L. Argerich and R. Wachenchauzer, "Variations of the Similarity Function of TextRank for Automated Summarization," *arXiv*, pp. 1-3, 2016.

[8] C. Mallick, A. K. Das, M. Dutta, A. K. Das and A. Sarkar, "Graph-Based Text Summarization Using Modified TextRank," *researchgate*, pp. 1-4, 2019.

[9] G. E. Radev and D. R., "LexRank:Graph-based Lexical Centrality as Salience in Text Summarization," *jair*, pp. 1-5, 2004.

[10] T. Uçkan and A. Karcı, "Extractive multi-document text summarization based on graph," *Egyptian Informatics*, pp. 1-6, 2020.

[11] M. G. Ozsoy, F. N. Alpaslan and I. Cicekli, "Text summarization using Latent Semantic Analysis," *SAGE*, pp. 1-3, 2011.

[12] Y. Liu, "Fine-tune BERT for Extractive Summarization," *arxiv*, pp. 1-5, 2019.

[13] B. Tan and V. Kieuvongngam, "Automatic Text Summarization of COVID-19 Medical Research Articles using BERT and GPT-2," *arxiv*, pp. 1-6, 2020.

[14] U. Khandelwal, K. Clark, D. Jurafsky and Ł. Kaiser, "Sample Efficient Text Summarization Using a Single Pre-Trained," *arxiv*, pp. 1-2, 2019.

[15] E. ShafieiBavani, M. Ebrahimi, R. Wong and F. Chen, "A Semantically Motivated Approach to Compute ROUGE Scores," *arxiv,* pp. 1-3, 2017.

[16] K. Ganesan, "ROUGE 2.0: Updated and Improved Measures for Evaluation of Summarization Tasks," *arxiv,* pp. 1-3, 2018.

[17] T. Zhang, V. Kishore, F. Wu, K. Q. Weinberger and Y. Artzi, "BERTSCORE: EVALUATING TEXT GENERATION WITH BERT," *arxiv,* pp. 1-3, 2020.