



REVA
UNIVERSITY

Bengaluru, India

Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

REVA Academy for Corporate Excellence (RACE)

Automated Literature Review with Transformers

Sanjeev Kumar Jha

SRN: R19MBA06

Date: 27th August 2022

MBA in Business Analytics

Capstone Project Presentation
Year: II

race.reva.edu.in



01 Introduction

Back Ground | Current status | Why this study

02 Literature Review

Seminal works | Summary | Research Gap

03 Problem Statement

Business Problem | Analytics Solution

04 Project Objectives

Primary & Secondary Objectives | Expected Outcome

05 Project Methodology

Conceptual Framework | Research Design

06 Business Understanding

Business Context | Monetary Impact

07 Data Understanding

Data Collection | Variables

08 Data Preparation

Pre-processing | Process | Techniques

09 Descriptive Analytics

Univariate | Bivariate | Hypothesis

10 Modeling

Machine Learning | Model Evaluation | Insights

11 Model Deployment

Applications | Demo

12 Suggestions and Conclusions

Insights | Next Step | Future Scope

13 Annexure

References | Publications | Plagiarism Score

- Knowledge creation in the field of business research is increasing at an incredible rate, This makes it difficult to stay current and at the forefront of research, as well as to analyse the aggregate data in a certain field of business research.
- Traditional literature evaluations are frequently lacking in completeness and rigor, and are done haphazardly rather than according to a defined technique.
- In the field of literature review the issues regarding the quality and credibility of these sorts of evaluations might be raised.
- Despite the significance of doing Systematic Literature Reviews (SLRs) for identifying research gaps in different research fields, performing SLRs manually is a difficult and time-consuming procedure.
- Primary goal of this project is to extract the reference paper from same area of work and summarize it based on existing research papers and generate a summary of the text which prevents duplication.
- Other research can use it to understand past work as well current status on that topic easily.

Literature Review

Seminal works | Summary | Research Gap

Sr. No.	Title	Author	Summary
1	Organizing Knowledge Syntheses. Knowledge in Society	Cooper, H. M. (2009)	A complete summary and critical analysis of the existing research and non-research literature on the subject at hand constitutes in a literature review. In their work author highlighted the importance of literature review.
2	Undertaking a literature review: a step-by-step approach	Cronin, P., Ryan, F., & Coughlan, M. (2008)	The objective of this is to keep the reader up to date with current literature on a certain topic and to serve as a foundation for another goal. A competent literature review collects information about a certain topic from a variety of sources.
3	Graph-Based text summarization using text rank.	Mallick, C., Das, A. K., Dutta, M., Das, A. K., & Sarkar, A. (2019).	It is a graph-based ranking model used for text processing. It is highly useful when extracting sentences and keywords by using an unsupervised approach. Text Rank is an extractive summarization technique. It is based on the concept that words which occur more frequently are significant.
4	Sample efficient text summarization using a single pre-trained.	Khandelwal, U., Clark, K., Jurafsky, D., & Kaiser, Ł. (2019).	Generative Pre-trained Transformer 2 (GPT-2) is an open-source artificial intelligence created by Open AI in February 2019. GPT-2 is a seq2seq model, it can also be customizable to perform the task of text summarization. GPT-2 translates text, answers questions, summarizes passages and generates text output.
5	Fine-tune BERT for extractive summarization.	Liu, Y. (2019).	BERT is a transformer used to overcome the limitations of RNN and other neural networks as long-term dependencies. It is a naturally bidirectional model that is pre-trained. Which is in our case summarization.

Problem Statement

Business Problem | Analytics Solution

- Search complexity of the paper, This project is to provide solution of the literature reviews which is done by different researchers during their researches.
- Time to write the Literature review
- NLP and state of art Transformer techniques a summary of whole literature review of existing paper in the current field will be generated.
- Researcher can understand what all the work has already been done in that field. Same can be referenced in their research work with proper citations.
- Generating the Literature Review for the paper, which prevents duplication

Project Objectives

Primary & Secondary Objectives | Expected Outcome

- The objectives of this study are to summarize the research paper using different papers on the same topic.
- Literature reviews/Abstract 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/Transformer summarization techniques.
- 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.

Project Methodology

Conceptual Framework | Research Design

01

Text Extraction

In this process, we have to retrieve the relevant content from a knowledge base.(Max 15 Pdf research paper)

02

Text Preprocessing

NLP technique to preprocess the text and cleaning it

03

Text Summarizing

NLP technique to get the Title feature, Sentence Length, Sentence Position, Sentence Frequency, Total Score

06

Deployment

Produce meaningful phrases and sentences. That is in the form of natural language from internal representation.

05

Evolution

Summarized Text Evolutions using Bert Score, Rouge

04

Summarized Text

In Summarized text user will get the summary of text based on some predefined criteria. TextRank (Summarization by ratio, Summarization by word count), LexRank, Latent Semantic Analysis (LSA),Luhn, LSTM, BERT,GPT-2

Business Understanding

Business Impact | Challenges | Monetary Impact

- 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.
- 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 well-informed and erudite, then only they can do a good research and contribute to the society.
- Researcher can develop rough-cut literature review within short time and they can explore further to enhance it.
- This tool reducing time for literature review by half with important key words and key sentences for research area

In every study endeavour, 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).

In this process, the researcher has to retrieve the relevant content from other papers from related area of research and provide the original text for the pre-processing stage.

In this paper data have been taken from different paper on the topic of customer segmentation. The text taken from the literature review of those paper and cleaned with different NLP techniques like. Regular expression stop words removal, punctuation marks, non-English words, numbers etc.

Data Preparation

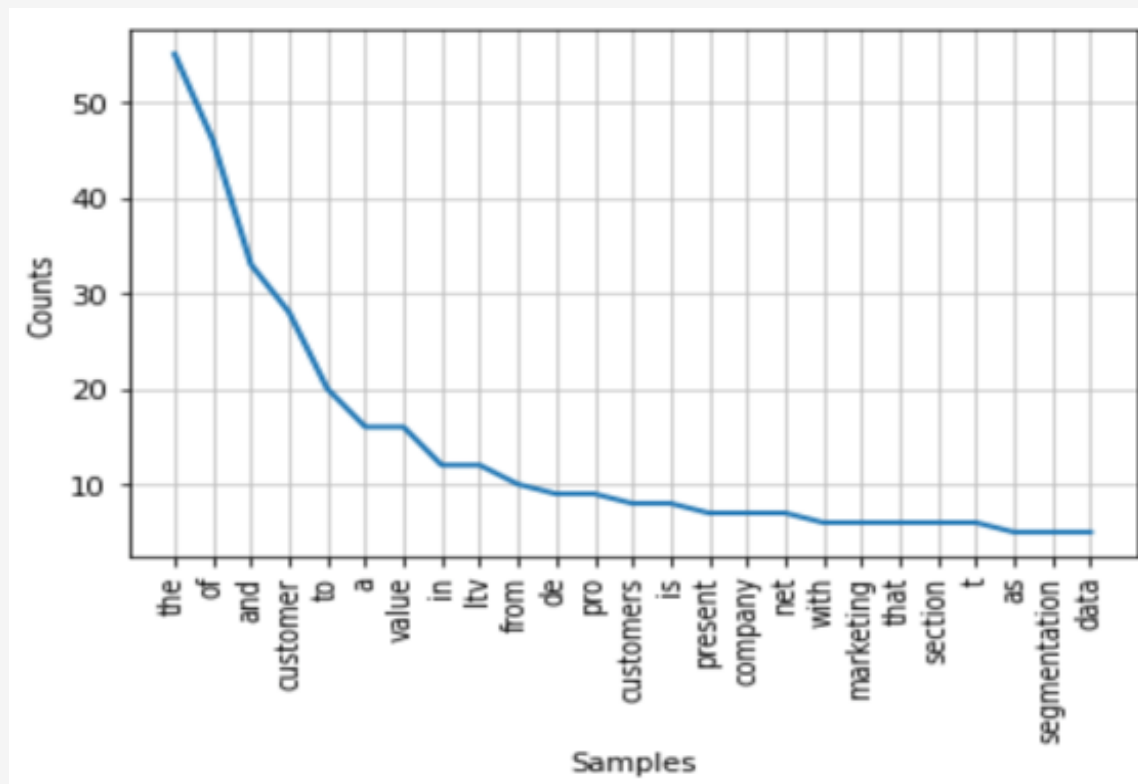
Pre-processing | Techniques

- 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.
- 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.

Customer relationship management (CRM) has 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 interest to invest in the development of customer acquisition, maintenance and development strategies. The business intelligence has an initial role to play in allowing companies to use technical expertise to gain better customer knowledge and Program for outreach by using clustering techniques to analyze customer attributes. Similar means are clustered together. Customer segmentation helps the marketing team to recognize and expose different customer segments that 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 purpose of building a good segmentation is to group customers into segments who will be similar and make the marketing team converge in an effective marketing plan. Clustering is an iterative process of knowledge discovery from vast amounts of raw and unorganized data. Clustering is a type of exploratory data mining that is used in many applications, such as machine learning, classification and pattern recognition. This paper is organized as follows. Section 1 reviews the previous studies related to customer value. This paper illustrates the limitations of existing studies and prepares the background reasons of this paper. Section 2 proposes a calculation model for measuring customer value applicable to a wireless telecommunication company. We apply real data of a wireless company to show the effectiveness of the proposed model. Section 3 compares the proposed customer value derived in Section 2 and proposes brief marketing strategies based upon the result of customer segmentation. Finally, Section 4 concludes this paper with the remark on the weaknesses of this study and future research directions. Related works. The definition of LTV and customer value has been studied under the name of LTV, Net Present Value (NPV), Customer Lifetime Value (CLV), and Customer Lifetime Profit (CLP). The definitions and calculations of LTV. The differences between the definitions are small. Table 1 shows the definitions of LTV. Considering the definitions above, we define 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 servicing customers, taking into account the time value of money. The horizontal axis represents the relationship stages, and the vertical axis denotes the type of relationship over time frame while vertical, type of customer value toward a company. A company forms various relationships according to the relationship stages—rudiment, beginning, fosterage, and expansion stage. A customer also gives a company various revenues, costs, and opportunities and potential benefits. Numerous studies have been conducted on LTV. For example, Bayo et al. [1] conducted a case study on LTV. In their researches, however, focused on Net Present Value (NPV) obtained from customers over the lifetime of transactions (Bayo et al., Gutsche, & Bauer, ; Berger & Nasr, ; Gupta & Lehmann, ; Roberts & Berger,). Dwyer et al. [2] tried to calculate LTV through modeling the retention and migration behavior of customers. Focused on making a decision of marketing investments, LaTour [3] proposed a method to calculate LTV. In his research, he used a method based on the approach. Hoekstra and Huizingh [4] also suggested a conceptual LTV model and categorized 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 model form based upon the profit deduction method is as follows. The profit is the difference between the revenue and the cost. The revenue is the sum of all future profits generated from a customer (Gupta and Lehmann). The net profit or loss to the firm from a customer over the entire life of transactions of that customer with the firm (Berger and Nasr). Expected profits from customers, exclusive of costs related to customer management (Blattberg and Deighton). The total discount value of the net profit from a customer (Berger and Nasr). The net present value of the customer (Pearson). The net present value of a future stream of contributions to overheads and profit (expected from the customer) (Jackson). The net present value of all future contributions to overhead and profit (Roberts and Berger). The net present value of the customer (Courtois et al.). Expert Systems with Applications (ESWA) [5].

Visualization of Data:With the help of word frequency graph and Word cloud , visualization of cleaned text has been provided.



Word frequency provides the number of times a word occurs.



Most popular words shown on word cloud

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.

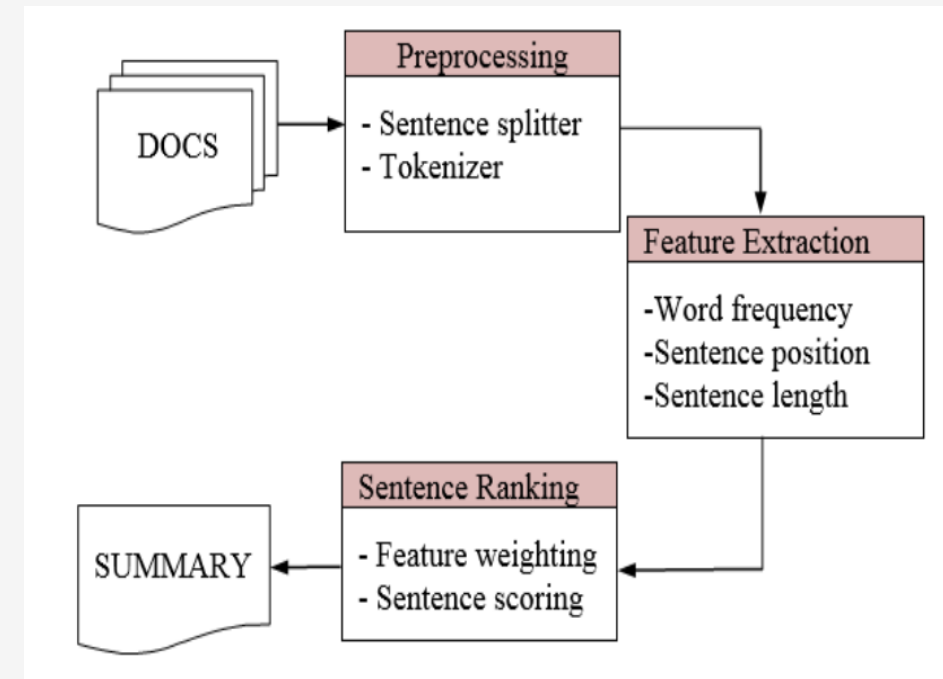
'Management and maintain of customer relationship have always played a vital role to provide business intelligence to organizations to build manage and develop valuable long term customer relationships. The importance of treating customers as an organizations main asset is increasing in value in present day and era. Organizations have an interest to invest in the development of customer acquisition maintenance and development strategies. The business intelligence has a vital role to play in allowing companies to use technical expertise to gain better customer knowledge and Programs for outreach. By using clustering techniques like k means customers with similar means are clustered together. Customer segmentation helps the marketing team to recognize and expose different customer segments that 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 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 and unorganized data. Clustering is a type of exploratory data mining that is used in many applications such as machine learning classification and pattern recognition. propose brief marketing strategies after segmenting customer base. This paper is organized as follows. Section reviews the previous studies related to customer value. This part illustrates 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 data 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 segmentation. Finally Section concludes this paper with the remark on the weaknesses of this study and future research directions. . Related works .. The definition of LTV Customer value has been studied under the name of LTV Customer Lifetime Value Customer Equity and Customer Profitability. The previous researches contain several definitions of LTV. The differences between the definitions are small. Table shows the definitions of LTV. Considering the definitions above we define 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 servicing customers taking into account the time value 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 relationships according to the relationship stages rudiment beginning fosterage and expiry stage. A customer also gives a company various revenues costs and opportunities and potential benefits. .. Models of LTV calculation There are a lot of researches on calculating customer value. The basic concept of these researches however focused on Net Present Value NPV obtained from customers over the lifetime of transactions Bayon Gutsche Bauer Berger Nasr Gupta Lehmann Roberts Berger . Dwyer tried to calculate LTV through modeling the retention and migration behavior of customers. Focused on making decision of marketing investment Hansotia and Rukstales suggested incremental value modeling using tree and regression based approach. Hoekstra and Huizingh also suggested a conceptual LTV model and categorized 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 model form based upon the proposed definition is as Fig. . The scope of CRM. Table Definitions of LTV Definition Article The present value of all future profits generated from a customer Gupta and Lehmann The net profit or loss to the firm from a customer over the entire life of transactions of that customer with the firm Berger and Nasr Expected profits from customers exclusive of costs related to customer management Blattberg and Deighton The total discounted net profit that a customer generates during her life on the household Bitran and Mondschein The net present value of the stream of contributions to profit that result from customer transactions and contacts with the company Pearson The net present value of a future stream of contributions to overheads and profit expected from the customer Jackson The net present value of all future contributions to overhead and profit Roberts and Berger The net present value of all future contributions to profit and overhead expected from the customer Courtheoux H. Hwang et al. Expert Systems with Applications '

Modeling Techniques | Modeling Process | Model Building

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 .

Summarization using TextRank

- TextRank is an extractive summarization technique.
- It is based on concept that **words which occur more frequently are significant.**
- Hence , the **sentences containing highly frequent words are important .**
- Based on this , the algorithm assigns scores to each sentence in the text .
- The top-ranked sentences make it to the summary.



Summarization using TextRank

'Section reviews the previous studies related to customer value. Section proposes a calculation model for measuring customer value applicable to a wireless telecommunication company. 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. Considering the definitions above we define 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 servicing customers taking into account the time value of money. .. Models of LTV calculation There are a lot of researches on calculating customer value. 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 Berger Nasr Gupta Lehmann Roberts Berger . The basic model form based upon the proposed definition is as Fig.'

- According to gensim source code, **at least 10 sentences** is recommend for the input
- No training** data or model building is required.
- It fits **not only English but also any other a bag of input** (Symbol, Japanese etc). You may also read TextRank [research paper](#) for detail understanding.
- From my experience, the result is **not good** in most of the time. It may due to **variety of words** and the **result is only a subset of input**.

Summarization using LexRank

- A sentence which is similar to many other sentences of the text has a high probability of being important.
- The approach of LexRank is that a particular sentence is **recommended by other similar sentences** and hence is ranked higher.
- Higher the rank, higher is the priority of being included in the summarized text.

'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 exploratory data mining that is used in many applications such as machine learning classification and pattern recognition. This paper is organized as follows. Section reviews the previous studies related to customer value. Section proposes a calculation model for measuring customer value applicable to a wireless telecommunication company. 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 definition of LTV Customer value has been studied under the name of LTV Customer Lifetime Value Customer Equity and Customer Profitability. Considering the definitions above we define 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 servicing customers taking into account the time value of money... Models of LTV calculation There are a lot of researches on calculating customer value. Table Definitions of LTV Definition Article The present value of all future profits generated from a customer Gupta and Lehmann The net profit or loss to the firm from a customer over the entire life of transactions of that customer with the firm Berger and Nasr Expected profits from customers exclusive of costs related to customer management Blattberg and Deighton The total discounted net profit that a customer generates during her life on the household list Bitran and Mondschein The net present value of the stream of contributions to profit that result from customer transactions and contacts with the company Pearson The net present value of a future stream of contributions to overheads and profit expected from the customer Jackson The net present value of all future contributions to overhead and profit Roberts and Berger The net present value of all future contributions to profit and overhead expected from the customer Courtho ux H. Hwang et al.'

Summarization using LSA (Latent Semantic Analysis)

- Latent Semantic Analysis is a unsupervised learning algorithm that can be used for extractive text summarization.
- It extracts semantically **significant sentences** to the matrix of term-document frequency.

'Customer segmentation helps the marketing team to recognize and expose different customer segments that think differently and follow different purchasing strategies. The main purpose of performing customer segmentation is to group people who have similar interest so that the marketing team can converge in an effective marketing plan. propose brief marketing strategies after segmenting customer base. Section reviews the previous studies related to customer value. We apply real data of a wireless company to the model in Section . The previous researches contain several definitions 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 researches 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.'

Modeling Techniques | Modeling Process | Model Building

Summarization using Luhn

- Luhn Summarization algorithm's approach is based on TF-IDF (Term Frequency-Inverse Document Frequency). It is useful when **very low frequent words as well as highly frequent words**(stopwords) are both not significant.
- Based on this, sentence scoring is carried out and the high-ranking sentences make it to the 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. Organizations have an interest to invest in the development of customer acquisition maintenance and development strategies. The main purpose of performing customer segmentation is to group people who have similar interest so that the marketing team can converge 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 definition of LTV Customer value has been studied under the name of LTV Customer Lifetime Value Customer Equity and Customer Profitability. Considering the definitions above we define 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 servicing customers taking into account the time value of money. The horizontal axis denotes the type of relationship over 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 Berger Nasr Gupta Lehmann Roberts Berger .Hoekstra and Huizingh also suggested a conceptual LTV model and categorized input data of the model into two types source of interaction data and time frame. Table Definitions of LTV Definition Article The present value of all future profits generated from a customer Gupta and Lehmann The net profit or loss to the firm from a customer over the entire life of transactions of that customer with the firm Berger and Nasr Expected profits from customers exclusive of costs related to customer management Blattberg and Deighton The total discounted net profit that a customer generates during her life on the household list Bitran and Mondschein The net present value of the stream of contributions to profit that result from customer transactions and contacts with the company Pearson The net present value of a future stream of contributions to overheads and profit expected from the customer Jackson The net present value of all future contributions to overhead and profit Roberts and Berger The net present value of all future contributions to profit and overhead expected from the customer Courtheoux H. Hwang et al.'

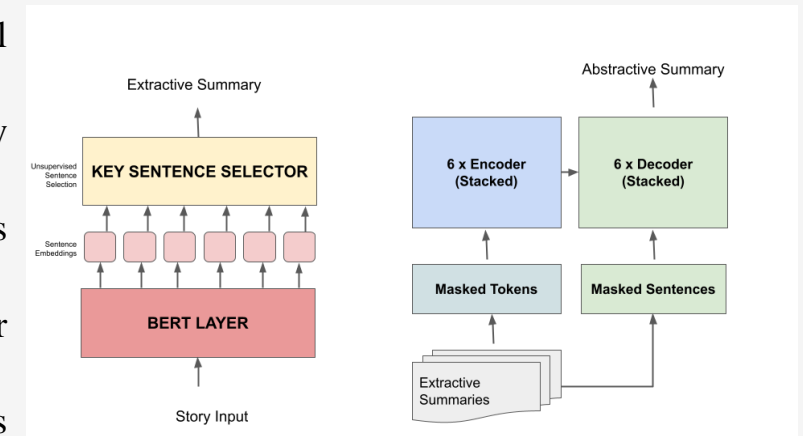
Summarization using KL-Sum

- Another extractive method is the KL-Sum algorithm.
- It selects **sentences based on similarity of word distribution as the original text.**
- It aims to lower the KL-divergence criteria (learn more).
- It uses greedy optimization approach and keeps adding sentences till the KL-divergence decreases.

'This part illustrates the limitations of existing studies and prepares the background reasons of this paper. Finally Section concludes this paper with the remark on the weaknesses of this study and future research directions..Related works ...The horizontal axis denotes the type of relationship over time frame while vertical type of customer value toward a company.A company forms various relationships according to the relationship stages rudiment beginning fosterage and expiry stage.A customer also gives a company various revenues costs and opportunities and potential benefits.The scope of CRM.Table Definitions of LTV Definition Article The present value of all future profits generated from a customer Gupta and Lehmann The net profit or loss to the firm from a customer over the entire life of transactions of that customer with the firm Berger and Nasr Expected profits from customers exclusive of costs related to customer management Blattberg and Deighton The total discounted net profit that a customer generates during her life on the house list Bitran and Mondschein The net present value of the stream of contributions to profit that result from customer transactions and contacts with the company Pearson The net present value of a future stream of contributions to overheads and profit expected from the customer Jackson The net present value of all future contributions to overhead and profit Roberts and Berger The net present value of all future contributions to profit and overhead expected from the customer Courtheoux H. Hwang et al.'

Summarization using BERT

- **BERT** is worked based on transformer used to overcome the limitations of RNN and other neural networks as long term dependencies.
- It is a pre-trained model that is naturally bidirectional. This pre-trained model can be tuned to easily perform the NLP tasks as specified, **Summarization** in our case.
- Summarization aims to condense a document into a shorter version while preserving most of its meaning.
- The BERT model used for unsupervised extractive summarization is a pre-trained transformer encoder model .
- BERT is not built to perform language generative task, its use for abstractive summarization is limited.

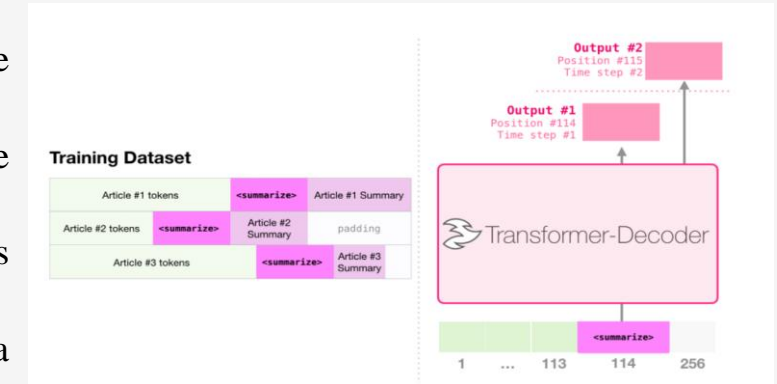


Bert Architecture

'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 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 segmentation. Related works .. The definition of LTV Customer value has been studied under the name of LTV Customer Lifetime Value Customer Equity and Customer Profitability. The previous researches contain several definitions of LTV. The horizontal axis denotes the type of relationship over time frame while vertical type of customer value toward a company.'

Summarization using GPT-2

- 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*”.
- Sequence-to-sequence (seq2seq) models based on the transformer encoder decoder architecture has been widely used for abstractive summarization.
- Abstractive summarization task requires language generation capabilities to create summaries containing novel words and phrases not featured in the source document.
- GPT2 to perform mapping from a selected keywords to a summary text, hence generating a summary abstractly.



Gpt-2 Architecture

'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 performing 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 and unorganized data. The previous researches contain several definitions of LTV. The basic concept of these researches 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 and migration behavior of customers. The basic model form based upon the proposed definition is as Fig. .'

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

Text Rank:

Bert Score

	Score
Bert	0.22

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

Latent Semantic Analysis:

Score

	Score
Bert	0.23

	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

Lex Rank:

Score

	Score
Bert	0.24

	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

Model Evaluation

Results | Interpretation | Insights

Luhn:

Bert Score

	Score
Bert	0.24

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

KL-Sum:

Score

	Score
Bert	-0.03

	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

BERT :

Score

	Score
Bert	0.42

	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

GPT-2 :

Score

	Score
Bert	0.28

	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

Model Deployment

Demonstration

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.

Upload File

Please upload PDF file

File is successfully uploaded.

Upload a File 1:

Choose File No file chosen

Upload a File 2:

Choose File No file chosen

Upload a File 3:

Choose File No file chosen

Upload a File 4:

Choose File No file chosen

Upload a File 5:

Choose File No file chosen

Upload

'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**']

Results and Insights

Key Findings | Suggestions

- 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 researcher 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.

Conclusion and Future Work

Proposed solutions | Scope for future work

- The literature review/abstract 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., & Wachenchauser, 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.
- 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 TextRank. *researchgate*, 1-4.
- Miller, D., & Atlanta, G. (2019). Leveraging BERT for Extractive Text Summarization on Lectures. *arxiv*, 1-6.



Additional Information | Plagiarism score

Automated Literature Review with Transformers

ORIGINALITY REPORT

13%

SIMILARITY INDEX

8%

INTERNET SOURCES

1%

PUBLICATIONS

9%

STUDENT PAPERS

PRIMARY SOURCES

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

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

libguides.usc.edu

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

Exclude quotes On

Exclude matches < 10 words

Exclude bibliography On



REVA
UNIVERSITY

Bengaluru, India

Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Annexure

Publications | Conferences





REVA
UNIVERSITY

Bengaluru, India

Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

*Special thanks to Dr. J. B. Simha and Dr. Shinu Abhi for
their guidance and all the support on this initiative.*



*Thank
you!*