International Journal of Engineering and Techniques - Volume 10 Issue 2, March 2024

LEGAL CASE DOCUMENT SUMMARIZATION USING AI

Vaishnavi Suryavanshi, Disha Naikwadi, Prof.Sneha Patil

Computer Department,Smt Kashibai navale College of Engineering,Pune.
vaishnaviis1010@gmail.com
Computer Department,Smt Kashibai navale College of Engineering,Pune.
dishadeepa15@gmail.com
Computer Department,Smt Kashibai navale College of Engineering,Pune.

Abstract:

Automatic Legal Document Summarization is an automated text summarization tool which is generated by a computer program. This projects aim to generate a suitable summary from a legal Documents for example affidavit, loan agreement & consulting Agreement. With the development of this module, it is hoped that this will decrease the time required for handling tender process, eliminating the need on using manual summarizingand providing an easy viewing for user. This program will be developed by incorporating Artificial Intelligence field of Natural Language Processing (NLP) techniques and also finding the most suitable methodology to handle a project development that deals on text summarization processes. Therefore a custom-made methodology are implemented which are based on SDLC methodology and a summarization process. In incorporating NLP technique, based on the existing summarization system technique on word counting and clue phrases for topic identification and word clustering are used for better interpretation of information. Apart from using NLP, other techniques such as theme extraction are also taken into consideration for better generation of the summary based Legal document summarization is an automated text summarization system that aims to generate relevant summaries from legal tender documents. The purpose of this paper is to explore the benefits of using automated text summarization in the legal tender process and showcase how incorporating Artificial Intelligence (AI) and Natural Language Processing (NLP) techniques can improve the efficiency of handling tender processes.

Keywords: Natural Language Processing, Software Development Lifecycle, Text Summarization, Artificial

1.INTRODUCTION

The growing volume of legal documents, including contracts, agreements, and other related materials, poses a significant challenge for legal professionals who need to review these documents within a limited timeframe. Manual summarization processes are time-consuming and prone to errors, making it difficult for legal practitioners to efficiently extract key information from extensive legal texts.This paper proposes a solution by leveraging AI and NLP techniques to develop an automated text summarization system specifically designed for legal tender documents. By implementing this system, we aim to reduce the time required for handling tender processes while ensuring that users have easy access to essential information through concise summaries.The Aim of a summary is to give the reader

2001). In this research, we focused on a problem referred to as legal text summarization. As ever larger amounts of legal documents become available digitally, interest in automated summarization has continued to grow in recent years. In this paper, we present our method for producing a very short text from a long legal document like Affidavit & Loan Agreement (a record of the proceedings of federal courts in India and Other United Sates) and present it as a Table(Rows & Column) style summary. The goal of this project is to develop a system to create a summary for the needs of lawyers, judges and experts in the legal domain. Our approach investigates the extraction of the most important units based on the identification of thematic structures of the document and the determination of semantic roles of the textual units

2.THISIS STATEMENT

This paper argues that incorporating AI and NLP techniques in the development of a custom-made methodology for legal document summarization can significantly enhance the efficiency of handling tender processes.

3.System Architecture

1. **Data pre-processing:** This stage involves cleaning and preparing the legal case documents for summarization. This may include removing irrelevant information, such as headers, footers, and citations, as well as identifying and normalizing legal terms and jargon.
2. **Summarization:** This stage involves applying a summarization algorithm to the pre-processed legal case documents. There are two main types of summarization algorithms: extractive and abstractive. Extractive algorithms identify and extract the most important sentences or phrases from the document, while abstractive algorithms generate new sentences that capture the essence of the document.
3. **Evaluation:** This stage involves evaluating the performance of the summarization system. There are a number of different evaluation metrics that can be used, such as ROUGE, BLEU, and F1 score.
4. **Rhetorical Role Classification:**
   - Aims to identify the function of a sentence phrase within a larger text.
   - Categorizes text elements based on their contribution to the overall argument or narrative.
   - Common roles include:
     1) Claim: introducing a main point or asserting information
     2) Evidence: supporting a claim with facts or reasoning.
     3) Counterargument: addressing an opposing viewpoint.
     4) Conclusion: summarizing a point or drawing an inference.

5.Relevance Classification:

- Determines whether a piece of text is relevant to a specific topic or query.
- Judges the informational connection between text and a given context..
- Can be binary (relevant/not relevant) or multi-class (highly relevant, partially relevant, not relevant).

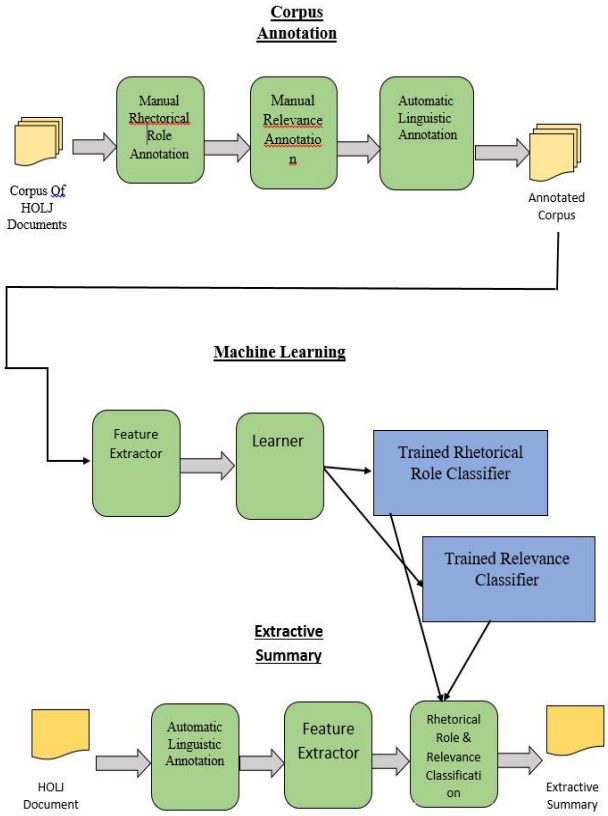


Fig.1: System Architecture Model

4. DESIGN AND ANALYSIS MODEL

The development of an automated text summarization system for legal tender documents

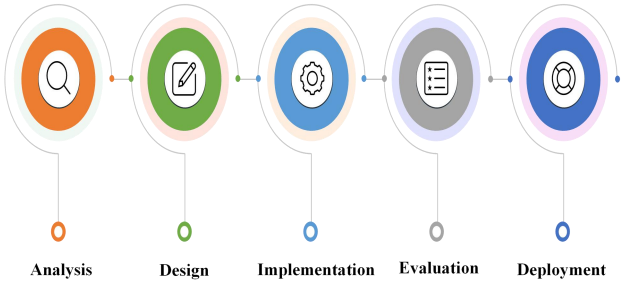


Fig.2:Agile Waterfall Model

1. **Requirements Analysis:** In this stage, we identify the specific needs and goals of the automated text summarization system for legal tender documents. This includes understanding the typical structure of these documents and determining what information should be extracted

2. **Design:** Following requirements analysis, we design the architecture of our system by defining the algorithms and techniques required for text summarization. This involves selecting appropriate AI and NLP models that can effectively process textual data while ensuring accuracy.

3. **Implementation:** The implementation stage involves coding the system according to the defined design specifications. We leverage existing libraries or develop custom algorithms as per our requirements.

4. **Evaluation:** To ensure that our automated text summarization system meets its intended purpose effectively, rigorous evaluation is required. This stage involves comparing the generated summaries with manually created summaries by legal professionals to measure accuracy and relevance.

5. **Deployment:** Once the system is evaluated successfully, it can be deployed in a production environment. Legal practitioners can then utilize this automated text summarization system to extract key information from legal tender documents efficiently analysis. In this paper, we propose a custom-made methodology based on the Software Development Life Cycle (SDLC) model(Agile Model).

5.CODE IMPLEMENTATION

5.1BART USED FOR SEQUENCE TO SEQUENCE SUMMARIZATION:

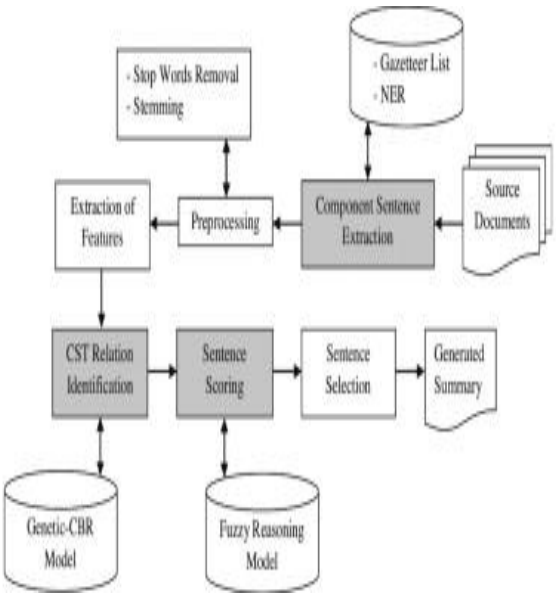


Fig.3:BART Model

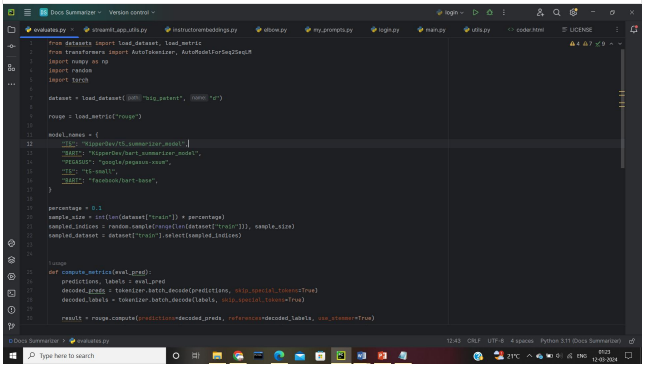
BART (Bidirectional and Auto-Regressive Transformers) is a transformer-based sequence-to-sequence model introduced by Lewis et al. (2020). BART can be effectively used for document summarization tasks.

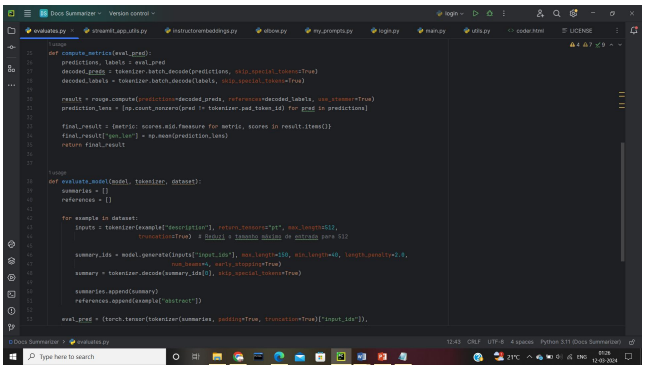
The BART model is pre-trained with a denoising autoencoding objective, where corrupted input is reconstructed by the model. This pre-training enables BART to perform well on various downstream tasks, including summarization. During fine-tuning for document summarization, BART is typically trained to generate concise summaries that capture the most important information from the input document.

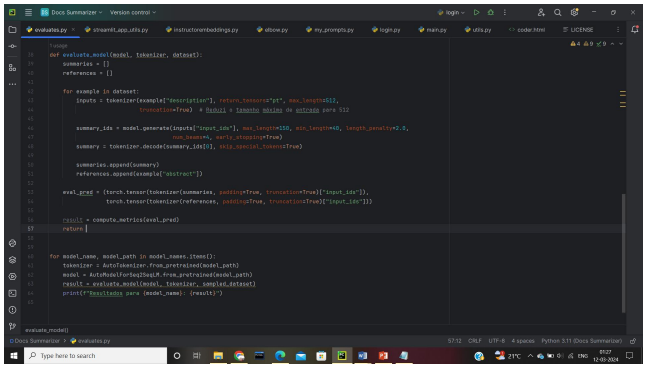
BART's ability to generate coherent and fluent summaries makes it suitable for document summarization tasks. It excels particularly in tasks where the input documents are longer and more complex, thanks to its bidirectional encoder and autoregressive decoder architecture.

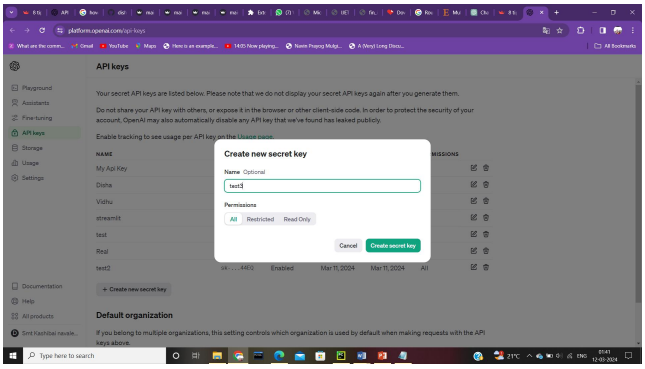
When evaluating BART for document summarization, you would typically use standard metrics such as ROUGE (Recall-Oriented Understudy for Gisting Evaluation) or BLEU (Bilingual Evaluation Understudy) to measure the quality of the generated summaries compared to human-written reference summaries.

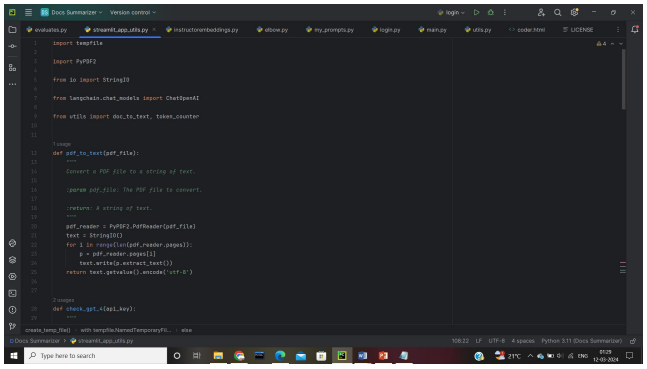
Overall, BART is a powerful model for document summarization tasks, and its performance can be fine-tuned and evaluated using standard evaluation metrics.

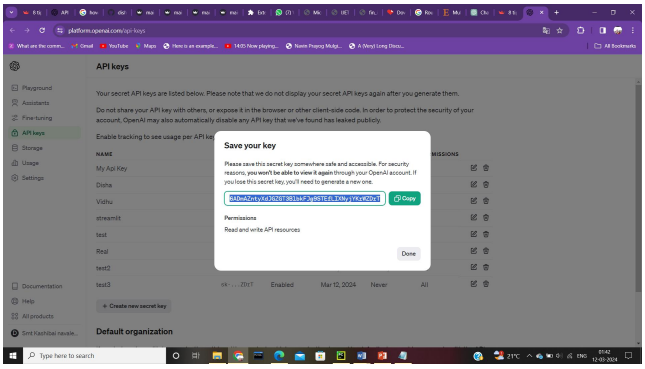






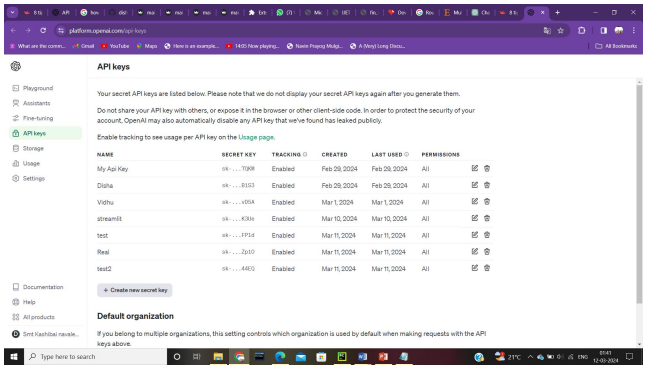




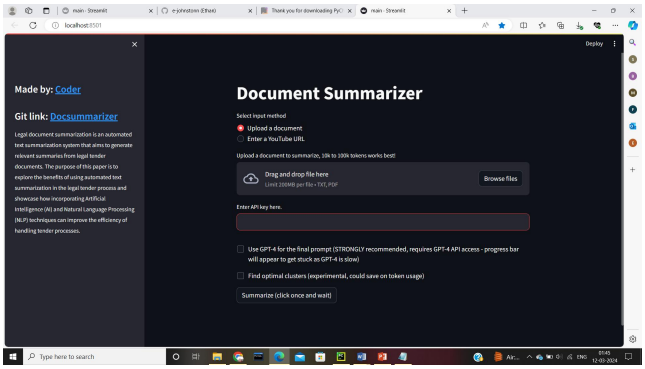


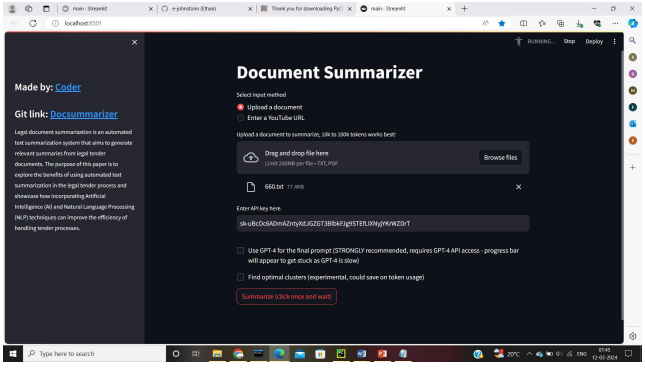
You need to create api key for summarization of legal documents. Streamlit application for the creation of api keys.
Using api key we load the documents and they can easily summarize the legal documents from the dataset.only text file and pdf's are allowed for document summarization.
Generating open AI API Key:

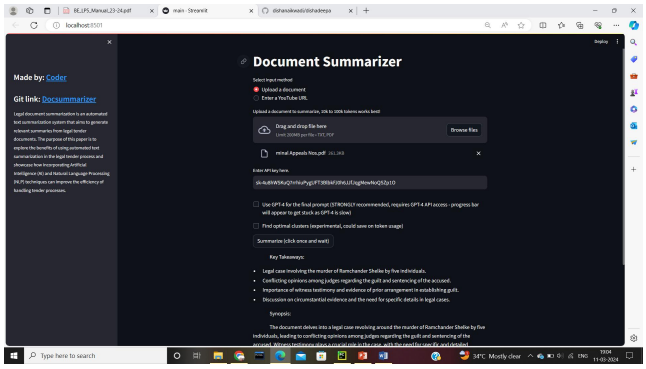
- Goto:https://platform.openai.com/account/api-keys
- Click on + create new secret key
- Enter an identifier name (optional) click on Create Secret key.

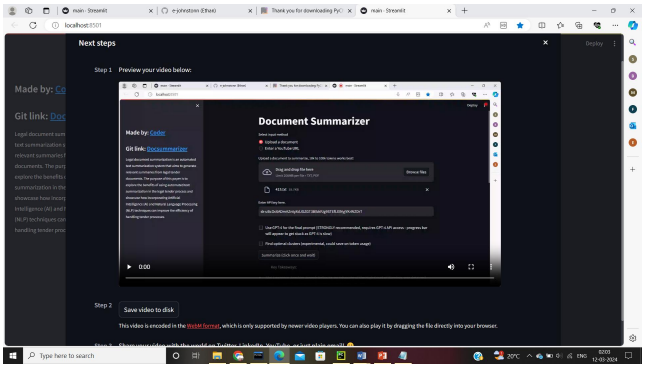


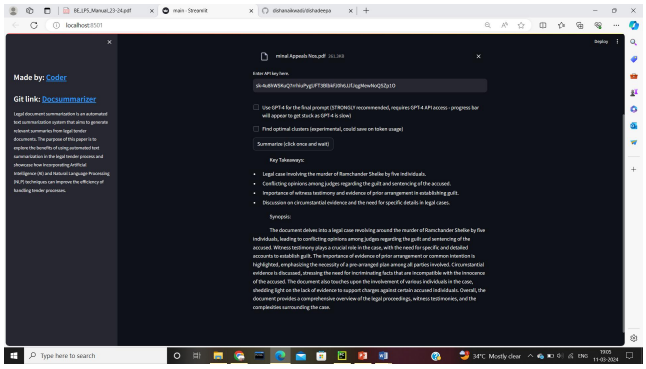
5.2GUI OUTPUTS:

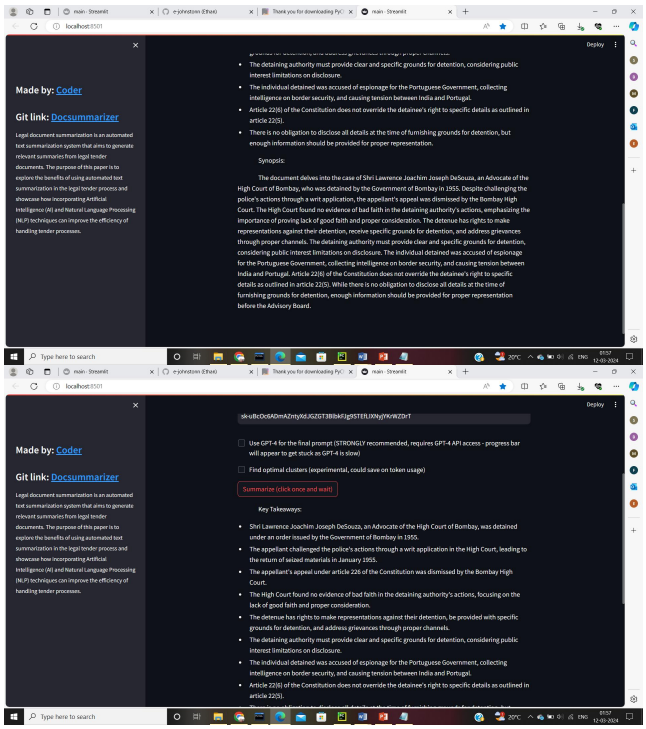












We also share the screen in any online meetings and also record the screen with audio. And we also save this record into our PC.This is Totally Open Source Application.

6.CONCLUSION

Incorporating AI and NLP techniques in the development of an automated text summarization system for legal tender documents offers significant benefits for efficient handling of tender processes. The proposed custom-made methodology based on SDLC provides a structured approach to ensure the successful implementation of such a system.

The use of AI-based approaches has already demonstrated promising results in other applications, such as virtual assistants and forensic data analysis. By leveraging these advancements and tailoring them to the specific requirements of legal document analysis, we can significantly reduce time consumption and improve user experience in handling extensive legal texts.

This paper highlights the importance of embracing technological advancements in the field of law to overcome challenges associated with manual document summarization. Through further research and development, automated text summarization systems have great potential to revolutionize how legal professionals handle large volumes of textual data efficiently.

7.ACKNOWLEDGEMENT

This document acknowledges the use of artificial intelligence (AI) for the summarization of legal documents related to Criminal Appeal 1970. The AI tool used is Streamlit Document Summarization AI document summarization system

The AI summary is intended to provide a general overview of the key points within the legal documents and expedite the review process. It should not be solely relied upon for complete comprehension or to replace a thorough legal analysis by qualified professionals.

We understand that the legal domain requires accuracy and interpretability. The summaries generated by the AI system have been reviewed by [Coder] to ensure they are consistent with the original documents and do not introduce misleading information.

8.REFERENCES

[1]Patil M.S, “A Hybrid Approach for Extractive Document Summarization Using Machine Learning and Clustering Technique”, (IJC SIT) International Journal of Computer Scienceand Information Technologies, Vol. 5 (2), PP: 1584-1586, 2014.

[2] Legal Document Summarization Using Nlp and Ml Techniques International Journal Of Engineering And Computer Science Volume 9 Issue 05 May 2020, Page No. 25039-25046

[3] Legal Document Summarization Using Nlp and Ml Techniques 1Rahul C Kore* , 2 Prachi Ray, 3 Priyanka Lade, 4Amit Nerurkar 1,2,3,4Vidyalankar Institute of Technology, Mumbai University, Mumbai, India Volume 09 Issue 05 May, 2020 Page No. 25039-25046

[4] A Survey of Legal Document Summarization Methods Sheetal Ajaykumar Takale Professor, Information Technology Department, VPKBIET, Baramati 0000-0002-6081-2524 International Journal of Advanced Research in Computer and Communication Vol. 12, Issue 7, July 2023 Page No.128-133

[5] M. Schraagen, F. Bex, N. van de Luijtgaarden, & D. Prijs (2022). Abstractive Summarization of Dutch Court Verdicts Using Sequence-to-sequence Models. In Proceedings of the Natural Legal Language Processing Workshop 2022 (pp. 76-87).

[6] S. Takale, S. Payal, S. Jagtap, P. Jagtap, and A. Khan, “Legal data assistive tool using deep-learning,” in 2023 Third International Conference on Secure Cyber Computing and Communication (ICSCCC), 2023, pp.791–796.

[7] A. Yadav, V.K. Harit, "Fault Identification in Sub-Station byUsing Neuro-Fuzzy Technique", International Journal of ScientificResearch in Computer Science and Engineering, Vol.4, Issue.6,pp.1-7, 2016

[8] Pramod Pardeshi, Ujwala Patil, "Fuzzy Association Rule Mining-A Survey", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.6, pp.13-18,2017