Low Level Design (LLD)

News Summarization

Written By	Diana Laveena DSouza
Document Version	0.1
Last Revised Date	05/12/-2022

Document Version Control

Change Record:

Version	Date	Author	Author
0.1	05/12/2022	Diana Laveena DSouza	Introduction & Architecture Defined

Reviews:

Version	Date	Reviewer	Comments
0.1	05/12/2022		Document Content, Version Control and Unit Test Cases to be added

Approval Status:

Version	Review Date	Reviewed By	Approved By	Comments

Contents

Abstract

Making news is hard enough, even if you don't think about tight deadlines and thorough fact-checking. A news piece must meet specific editorial criteria, such as accuracy, timeliness, and availability of sources. On its own - writing a news piece is not a big deal. But there is much stuff going on in the world. And the reality is that a news media platform must deliver news in time to remain competitive and engage with the target audience. Thus, Text Summarization is a cost-effective and time-saving option for the media and journalists. Think about it as a helping hand for the journalist.

1 Introduction

1.1 Why is Low-Level Design Document?

The goal of LLD or a low-level design document (LLDD) is to give the internal logical design of the actual program code for News Summarization. LLD describes the class diagrams with the methods and relations between classes and program specs. It describes the modules so the programmer can directly code the program from the document.

1.2 Scope

Low-level design (LLD) is a component-level design process that follows a step-by-step refinement process. This process can be used for designing data structures, required software architecture, source code and ultimately, performance algorithms. Overall, the data organization may be defined during requirement analysis and then refined during data design work.

2 Technical specifications

2.1 Dataset

DataSet	Finalized	Source
Cornell Dataset	yes	https://lil.nlp.cornell.edu/newsroom/

2.1.1 Dataset Overview

The Dataset consists of news articles and their summary.

There are 76,000 records in the training set and 70 in the validation set for both Extraction and Abstraction Summarization.

News articles and their summary

	text	summary
0	By MATT SCHWARTZ in Houston and WENDELL JAMIES...	Bleeding from a massive chest wound, Tejano st...
1	By HOLLY RAMER, Associated Press CONCORD, N.H....	By HOLLY RAMER, Associated Press CONCORD, N.H....
2	Men in battle-fatigues have raided the Moscow ...	Men in battle-fatigues have raided the Mo...
3	Like all gadgets, cellphones can break. In fac...	Simple home remedies for repairing your mobile...
4	I visited Saudi Arabia in September 2008, arri...	Hurricane Ike, Ramadan and the billionaire pri...

2.1.2 Input schema

Feature Name	Datatype	Size	Null/Required
text	text	1000	Not Required
summary	text	200	Not Required

2.2 Predicting Summary

- The system presents the set of inputs required from the user.
- The user gives the required information.
- The System should be able to do extractive and abstractive summarization.

2.3 Database

The database used in this project to store the data is the Cassandra database.

2.5 Deployment

3 Technology Stack

Front End	HTML/CSS
Backend	Python Flask
Database	Cassandra
Deployment	GCP

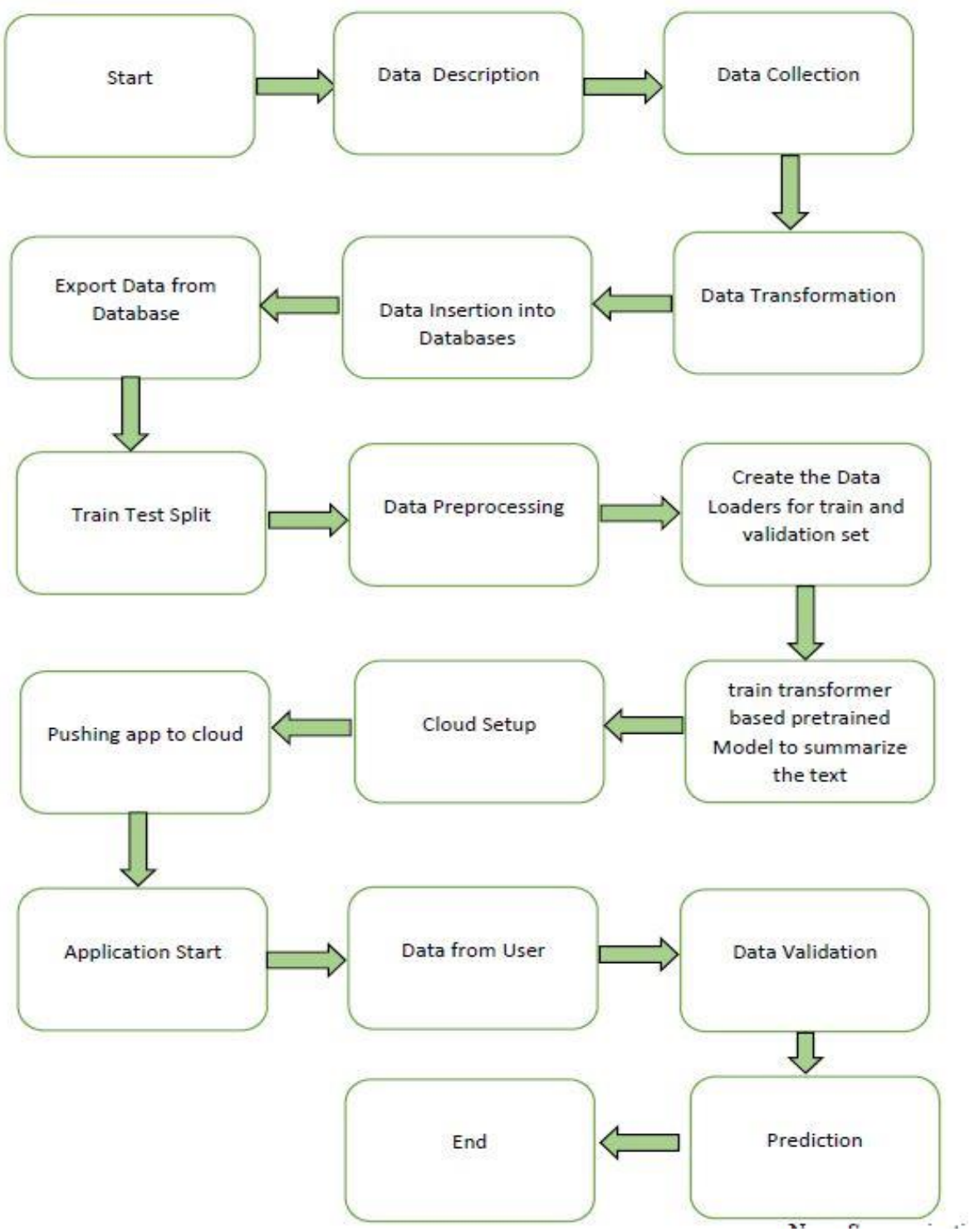
4 Proposed Solution

Refer: https://huggingface.co/course/chapter7/5?fw=pt

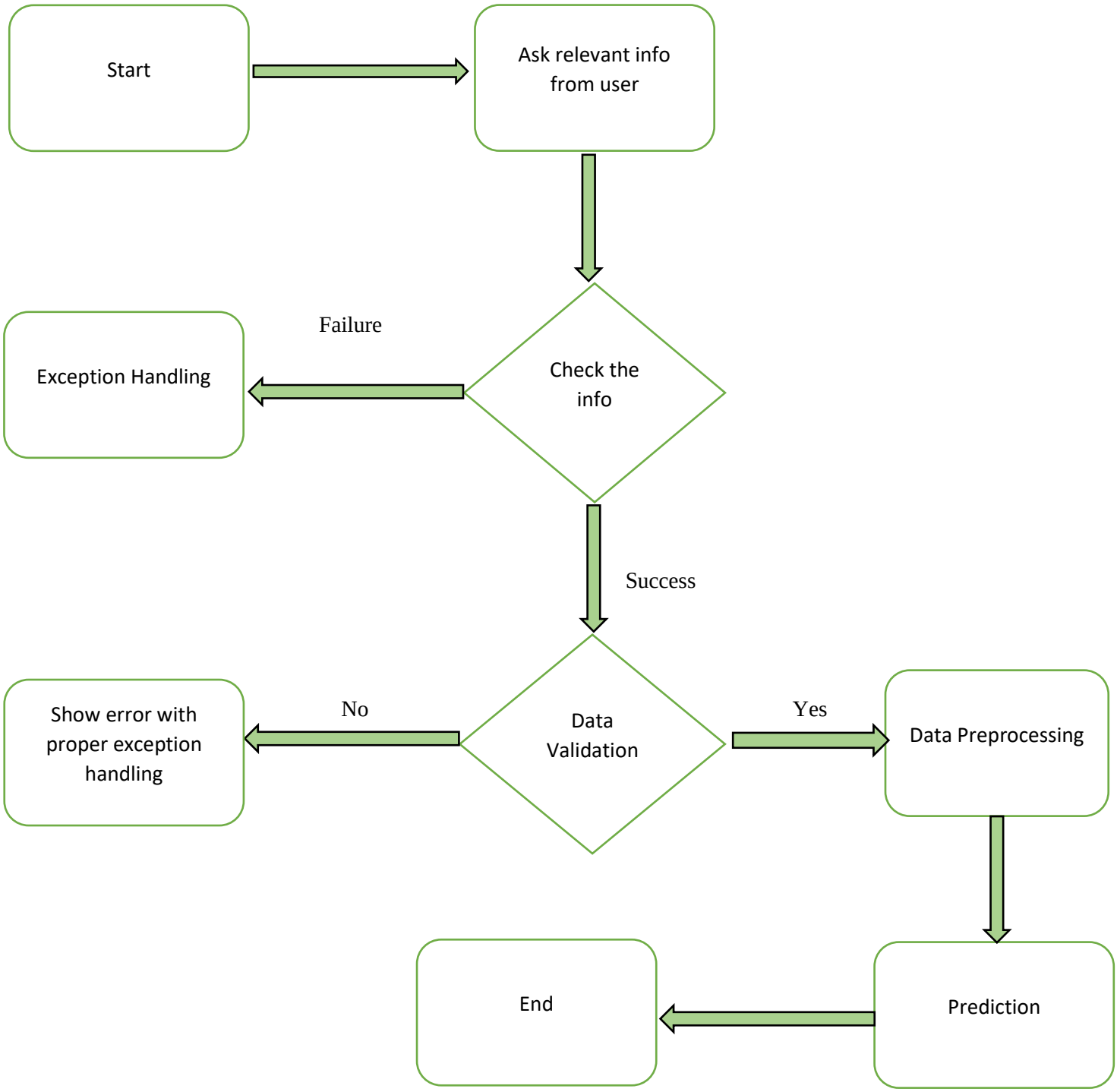
The research paper proposed various methods like GPT-2, PEGASUS, T5, mT5, BART, mBART-50 for text summarization. Finally, we selected the T5-base method.

Baseline Model: T5-Base Model

5 Model Training/Validation Workflow



6 User I/O workflow



7 Test Cases

Test Case Description	Pre-Requisite	Expected Result
Verify whether the Application URL is accessible to the user	1. Application URL should be defined	The application URL should be accessible to the user.
Verify whether the Application loads entirely for the user when the URL is accessed	1. Application URL is accessible 2. Application is deployed	The Application should load entirely for the user when the URL is accessed.
Verify whether the user can input the text in all input fields	1. Application is accessible	The user should be able to input the text in all input fields.
Verify whether the user gets Submit button to submit the inputs.		The user should get Submit button to submit the inputs.
Verify whether the user is presented with results on clicking submit.		The user should be presented with results on clicking submit