Requirement Analysis

Functional Requirement

The application should classify the article into a single topic from a group of 16 different topics based on its contents and there should be a graphical representation comparing the different algorithms applied for classification. There should also be a graphical representation showing how much an article is related to each topics.

Non-Functional Requirement

The application should be consistent at classification whenever same article is passed into the pipeline.

Feasibility Study

Feasibility studies aim to objectively and rationally uncover the strengths and weaknesses of an existing business or proposed venture, opportunities and threats as presented by the environment, the resources required to carry through, and ultimately the prospects for success. The feasibility analysis is divided into three parts as described below:

Technical Feasibility

The application does not need much technical expertise for usage. The project is developed is general use. In order to access this web application, the user needs a working browser. The main requirement of the system from a developer’s view is a web server capable of handling the content, Internet connection.

Economic Feasibility

The system is economically feasible as the only cost required will be the cost to host and run the website on a server and maintain the system.

Operational Feasibility

It is concerned with operating capabilities of the system. Since it is a web-based application, it is quite easy to handle the system with a normal web surfing skill. For the efficient operation, only a general purpose computer is required. And the user interface is friendly. Hence, the system is feasible operationally.

Methodology

Dataset Preparations

The Nepali language belongs to one of the most common scripts, Devanagari, invented by Brahmins around the 11th century. It consists of 36 consonant symbols, 12 vowel symbols and 10 numeral symbols along with different modifiers and half forms. According to current census research, 17 million people worldwide speak the Nepali language. Nepali language character set is given in Table 1.

A collection of Nepali news was collected from various online Nepali News portals using web crawler. The news portal namely ratopati.com, setopati.com, onlinekhabar.com, and ekantipur.com were used to gather text related to different news types. The distribution of news type in the Nepali news corpus is as shown in Table 2.

Preprocessing

The text preprocessing cleans the text data to make it ready to use in training and testing of the machine learning model. Preprocessing is done to reduce the noise in the text that helps to improve the performance of the classifier and speed up the classification process, thus aiding in real time news classification. The main preprocessing techniques used are given below.

1. Tokenization: Breakdowns the text into sentences.

and then words. Vertical bar, question mark, and full

stop are used to break down the sentences and while

space and comma are used to break down the words. 2. Special symbol and number removal: Special symbols like!, , , , , >, < \, ,@, #, $, %, ^, &. \*, ), (, , , , ~, ø, [, ], ', ", etc. and numbers, those