Identifying the Writing Style of Bangla Language Using Natural Language Processing

Introduction:

Bangla is one of the 8th major spoken languages around the world and like other widely spoken languages, it is a very morphologically rich language. It has two styles, one is the standard literary style, known as Sadhu Bhasha and the other one is a standard colloquial style which is known as Cholito Bhasha. Mixing both styles in a written document is considered a grammatical error in the Bangla language known as Guruchondali Dosh. This research aims to develop an algorithm to identify the style of a Bangla paragraph i.e. whether it is in Sadhu Bhasha or Cholito Bhasha from a given Bangla paragraph input. It's a contribution towards finding the Goruchondali Dosh which is a common grammatical mistake in the written Bangla language as it was observed that several research works for identifying Bangla grammar mistakes is not so notable whereas it is a common trend in other language researchers.

Literature Review:

- The literature review focuses on syntax and morphology differences in natural languages.
- Emphasizes slight grammar variations to make languages distinct; cites American and British English.
- Grammar, defined as syntax and morphology combinations, is crucial for forming sentences.

Methodology:

- The research methodology section focuses on the theoretical aspects of the study.
- Discusses research subject and instrumentation; "instrument" refers to tools like questionnaires.
- Two types of data were collected: Sadhu data and Cholito data.

Result Analysis:

- Cholito data comprises 148,702 sentences, ranging from 24 to 5,180 sentences per paragraph.
- The frequency graph shows an optimal threshold between 72.5% and 77.5% for identifying Sadhu and Cholito styles.
- Experimental results indicate an overall accuracy of 75%, derived as a mean value from both rates.

Conclusion:

- The research involved developing a library for Sadhu word extension in the Bangla language.
- Enables identification of Sadhu or Cholito style.
- Suggests potential future extensions, including style conversion tools for users to switch styles.
- Highlights broader applications of Natural Language Processing (NLP) in collaboration with IoT, Security, and Data Mining.