Task 4: Adapt Harmony to other Domains

Team :

Saurabh Jaiswar

Ashraf Ansari

Sushma Shambhu

Shubham Dutta

Akhil Chopra

OBJECTIVE

The primary objective of our group task is to enhance the efficiency and usability of the Harmony dataset by implementing data dimensionality reduction techniques. Specifically, our goals are to:

1. **Increase Data Efficiency**: Optimise the dataset by transforming various time frames into a consistent format, making it easier to analyse and interpret.

2. **Reduce Data Dimensionali**ty : Apply dimensionality reduction methods to minimise the complexity of the dataset, thus facilitating quicker data processing and more effective analysis.

3. **Enhance Data Consistency:** Ensure that all temporal data points are converted into a uniform unit (days), thereby standardising the dataset and eliminating inconsistencies.

4. **Improve Data Quality:** Identify and address entries that are vague or complex, converting them into precise numerical values where possible to enhance data quality and reliability.

5. **Facilitate Data Analysis:** By reducing the dimensionality of the data and improving its consistency, we aim to make the dataset more manageable and easier to work with for further statistical analysis, machine learning, and other research purposes.

6. **Support Decision Making:** Provide a cleaner, more efficient dataset that can be utilised for more accurate and reliable decision-making processes in subsequent research or operational tasks.

By achieving these objectives, our task will result in a streamlined clinical dataset that is both efficient and effective for various analytical purposes, supporting better data-driven decisions and insights.

METHODOLOGY USED

The methodology employed for this study consisted of utilising the Sentence Transformer model developed by Hugging Face. This model employs a deep learning architecture to encode textual inputs into fixed-dimensional vectors, enabling the generation of semantically meaningful representations. Specifically, we leveraged pre-trained transformer-based architectures such as BERT, RoBERTa, and DistilBERT, fine-tuned on specific downstream tasks or datasets when applicable. The Sentence Transformer model facilitated various natural language processing (NLP) tasks, including but not limited to text similarity computation, text classification, and text generation. By harnessing the power of transfer learning and transformer-based architectures, we aimed to enhance the performance of NLP tasks within the scope of our study

CONCLUSION

In conclusion, our group task aimed to enhance the efficiency and usability of the Harmony dataset through a comprehensive approach focused on data dimensionality reduction techniques. Through our concerted efforts, we successfully achieved our predefined objectives

In summary, our efforts have culminated in a streamlined clinical dataset that is both efficient and effective for a wide range of analytical purposes. By supporting better data-driven decisions and insights, our work contributes significantly to advancing research and operational endeavours in the relevant domains.