CS4120 Final Project Proposal - Draft

Group Members: Jai Gollapudi, Garvin Chan, Aden Lu

Candidate Project 1: Sentiment Analysis of Conversations with Older Adults

1. What do you want to do?
Develop and evaluate sentiment analysis models to assess the emotional content of conversations with older adults to gain insights into their emotional well-being.

2. What data would you need?
We would need transcripts of conversations with older adults from our service learning sessions (if possible with proper consent and anonymization). Additionally, we would need emotion-labeled datasets like ISEAR or EmotionLines for training and benchmarking.

3. What tools would you be using?
We would be using ML/NLP frameworks such as Scikit-learn for traditional classifiers, TensorFlow or PyTorch for neural networks, and NLTK and spaCy for preprocessing and tokenization.

4. What is a list of few candidate models you could use?
We could use a variety of models such as the VADER Sentiment Analyzer (Rule-Based Model), Logistic Regression or Support Vector Machine (Traditional ML Model), or Fine-tuned BERT/RoBERTa models (Deep Learning Model).

5. Any other resources that you'll need to use?
We will need access to GPUs via Google Colab and consent forms and privacy protocols for data collection.

6. What would be your desired end results?
We would hope to achieve a comparative analysis of different sentiment analysis models on the task, be able to identify the most effective model for analyzing elderly conversations, and gain insight into the prevalent emotional states of participants.

Candidate Project 2: Topic Modeling of Conversations with Older Adults

1. What would you want to do?
Implement and compare topic modeling techniques to identify common themes in conversations with older adults to better understand their interests and needs.

2. What data would you need?
Transcripts of conversations with older adults from our service learning sessions (if possible with proper consent and anonymization) and publicly available datasets like Alzheimer's Society Dementia Talk.

3. What tools would you be using?

We would be using Scikit-learn, NLTK, spaCy, Gensim for LDA, Sk-learn for NMF, pyLDAvis, and matplotlib

4. What is a list of few candidate models you could use?
Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF), and BERTopic for transformer-based topic modeling.

5. Any other resources that you'll need to use?
Stopwords lists (excludes words common in elderly speech but not meaningful for topics), computational resources (access to GPUs via Colab), and ethical guidelines (consent forms and privacy protocols).

6. What would be your desired end results?
A set of interpretable topics that reflect the main themes in conversations, an evaluation of which topic modeling approach yields the most meaningful results, and recommendations for program activities based on identified interests.

Candidate Project 3: Text Summarization for Memory Support in Older Adults

1. What would you want to do?
Develop and evaluate text summarization models to create concise summaries of conversations or texts, aiding older adults with memory support and information retention.

2. What data would you need?
Transcripts of conversations with older adults from our service learning sessions (if possible with proper consent and anonymization) and dialogue datasets like SAMSum Corpus or AMI Meeting Corpus.

3. What tools would you be using?
Scikit-learn for traditional classifiers, TensorFlow or PyTorch for neural networks, and NLTK and spaCy for preprocessing and tokenization. Additionally, we could use models like VADER Sentiment Analyzer and Hugging Face Transformers for pre-trained models like BERT.

4. What is a list of few candidate models you could use?
Frequency-Based Sentenced Scoring, TextRank Algorithm/Lead-3 Baseline, and BART/T5 (baseline models).

5. Any other resources that you'll need to use?
Access to GPUs via Google Colab (computational resources) and consent forms and privacy protocols for data collection (ethical guidelines).

6. What would be your desired end results?
A functioning text summarization tool that can generate concise and coherent summaries of conversations or texts relevant to older adults and insights into which summarization techniques are most beneficial for aiding older adults in recalling information.