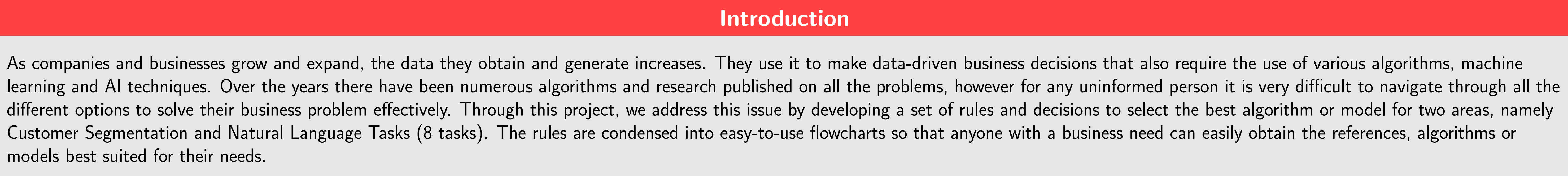
¹IIT Madras

Results

All the obtained results, documentation and code can be found at https://github.com/KausikN/MTech_MTP_UserSegmentation and https://github.com/KausikN/MTech_MTP_TextProblems. All the flowcharts for language tasks follow the same structure and hence only the flowchart of sentiment analysis is shown.

Customer Segmentation Flowchart

-
- ```

graph TD
 Start([Start]) --> TrueUnknown{True customer segments not known
(Problem: Clustering)}
 TrueUnknown --> NumKnown{Number of Segments Known?}

 NumKnown -- Yes --> NeedSoftYes{Need Soft Clustering?}
 NumKnown -- No --> NeedSoftNo{Need Soft Clustering?}

 NeedSoftYes -- Yes --> T1B_Yes[T1_B]
 T1B_Yes --> GMU[GMU (Variants)]
 GMU --> FuzzyCMeans[Fuzzy CMeans]
 FuzzyCMeans --> T1B_Yes
 GMU --> NotWorkingYes[Not working / Need better segments?]
 NotWorkingYes --> T1C[Complex Combined Algorithms]
 T1C --> SelfOrganizingMap[Self-Organizing-Map - kMeans++]
 SelfOrganizingMap --> Steps[Steps:
1. Train SOM of large enough size on the dataset
2. Perform kMeans++ on weight vectors of trained SOM to arrive at final cluster centers
3. For any data point, compute its winning vector in SOM and select the closest cluster center computed in kMeans to get its assigned final cluster]
 Steps --> T1C

 NeedSoftYes -- No --> T1A_Yes[T1_A]
 T1A_Yes --> CategoricalFeatures{Categorical Features Present?}
 CategoricalFeatures -- Yes --> KPrototypes[KPrototypes]
 CategoricalFeatures -- No --> NumSamples10K{NumSamples <10K?}
 NumSamples10K -- No --> MiniBatchKMeans[MiniBatch kMeans++]
 MiniBatchKMeans --> NotWorkingNo[Not working / Need better segments?]
 NotWorkingNo --> T1C
 NumSamples10K -- Yes --> AgglomerativeClustering[Agglomerative Clustering]
 AgglomerativeClustering --> Birch[BIRCH]
 Birch --> SpectralClustering[Spectral Clustering]
 SpectralClustering --> NotWorkingNo

 NeedSoftNo --> OutlierDetection{Outlier Detection Needed?}
 OutlierDetection -- No --> T1A_No[T1_A]
 T1A_No --> ElbowMethodNo[Elbow Method]
 ElbowMethodNo --> T1A_No

 OutlierDetection -- Yes --> TryInOrder[Try in order:
1. Mean-Shift
2. Affinity-Propagation
3. Spectral Clustering]
 TryInOrder --> NotWorkingTry[Not working / Need better segments?]
 NotWorkingTry --> T1C
 TryInOrder --> DBSCAN[DBSCAN (with OPTICS)]
 DBSCAN --> NotWorkingDB[Not working / Need better segments?]
 NotWorkingDB --> T1C
 DBSCAN --> T1D[T1_D]
 T1D --> T1C

```
- The flowchart illustrates the Adaptive Clustering Algorithm (ACA) for customer segmentation. It begins with a decision on whether true customer segments are known. If not, it checks if the number of segments is known. If yes, it proceeds to T1\_B, which uses GMU (Variants) and Fuzzy CMeans. If no, it checks if soft clustering is needed. If yes, it proceeds to T1\_A, which uses KPrototypes, MiniBatch kMeans++, or Agglomerative Clustering (BIRCH) followed by Spectral Clustering. If no, it checks for outlier detection. If needed, it tries a sequence of algorithms (Mean-Shift, Affinity-Propagation, Spectral Clustering) and then DBSCAN (with OPTICS). If not needed, it uses the Elbow Method. If the initial attempts fail, it proceeds to T1\_C, which uses Complex Combined Algorithms and Self-Organizing-Map - kMeans++ with specific steps for training and clustering. The final output is T1\_D.

## Sentiment Analysis Model Selection

