The submission directory contains the following files:

1) main_kmeans.py: Runs just the KMeans Algorithm over range of K [2, 6], and for each value of K, the algorithm is run 10 times, and the average Silhouette score is taken and displayed in the terminal. The optimal value of K is printed in the terminal, and the variations of the average silhouette score is plotted. Execution time: 1 - 2 mins. Commented in detail.
   Syntax to run: python main_kmeans.py

2) main_hierarchial.py: Runs just the Hierarchical Clustering Algorithm using single linkage over the dataset. Calculates the Silhouette score at each value of K [2, number of data samples]. Stores the calculated Silhouette scores in a csv file called “data.csv”. Also plots the variation of average Silhouette coefficient. Execution time: 5 - 6 mins. Commented in detail.
   Syntax to run: python main_hierarchial.py

3) main_comb.py: Main program that combines both KMeans clustering and Hierarchical Algorithm and does all the tasks mentioned in the report sequentially. First it runs the KMeans Algorithm (single iteration, not 10 times) over the range of K [2, 6] and then prints the SIlhouette scores for each value of K. Then it calculates the optimal values of K and runs the hierarchical algorithm to reach that value of K. Then, it computes the Jaccard similarity matrix and prints it. It also saves the generated clusters in the text files as mentioned in the report. Execution time: around 7 mins. Commented sparsely, please refer to comments in main_kmeans.py and main_hierarchial.py, since each individual part works in the same way
   Syntax to run: python main_comb.py

4) agglomerative.txt: The Main output text file containing the clusters generated by the Hierarchical Clustering algorithm, in the format as specified in the report

5) kmeans.txt: The Main output text file containing the clusters generated by the KMeans Algorithm, in the format as specified in the report.

6) cricket_4_unlabelled.csv: The csv file containing the dataset

7) scikit_visuals.py: Code for visualization of the data and performing KMeans Clustering by the Scikit library for comparison of performance

8) 5D Plot.html: Opens up an interactive visualization of the data in the web browser.

9) agglomerative_5.txt: Output text file for hierarchical clustering when the optimal K was given as 5 (to support the discussions made in the report)

10) kmeans_5.txt: Output text file for KMeans clustering when the optimal K was given as 5