Deliverable 1 (Revised Version)

Topic Modeling

1) **Dataset:** https://www.kaggle.com/code/richardnch/topic-modelling-on-neurips-papers/input. The dataset consists of almost 10,000 research papers submitted to the annual Conference on Neural Information Processing Systems (NeurIPS) between 1987 and 2019.

2) **Machine Learning Models:** Latent Dirichlet Allocation (LDA) and k-means clustering algorithm will both be trained on the NeurIPS dataset.
   Evaluation Metrics: For LDA: An intertopic distance map and a perplexity score will be used to evaluate how well the model assigns topics to a document.
   For k-means: Silhouette Score and Plot

3) **Application:** Make a poster presentation summarizing the theory behind both LDA and TF-IDF/k-means as well as the accuracy achieved by both models on the NeurIPS dataset.