From: **ZHOU Qiqi** qzhouam@connect.ust.hk
Subject: CSIC 5011: Project 1 Review
Date: 21 April 2023 at 11:15 AM
To: datascience.hw@gmail.com

ZQ

G13. **LI Haobo, CHEN Zixin, TENG fei, SHENG Rui**.
--- NIPS Conference Papers 1987-2015 Data Set.

- Summary of the report：
This project studies the frequency of word groups in NIPS conference papers. They utilize a framework for grouping words, predicting the evolution of word group frequency over time, and visualizing the evolution. Their analysis is based on a dataset of 5811 NIPS conference papers from 1987 to 2015, and examines the changes in topic and writing style within the NIPS community.

- The report has several strengths, including:

1. Visualization: This work provides a page for readers to inspect the result in an interactive way, not only the static image in the poster.
2. Motivation: The study of accepted papers is a good meta-problem to research how research itself is evolving and to gain a reflection on it.

- The weaknesses of the report:
1. Clarity: In the bubble figure, the meaning of 'size' and 'y-axis' is confusing as they appear to have the same meaning in the description, but are not consistent in the figure. Additionally, it is unclear which years the predicted results pertain to, and the MAE evaluation, as described in Section 3, is also not provided.

- Evaluation on Clarity and quality of writing (1-5): 4
Basically clear and organized, but some statements are missing. see weakness.

- Evaluation on Technical Quality (1-5): 5
Reproducible code is provided. Visualization is informative.

- Overall rating: 4.5
- Confidence on your assessment (1-3): 2

G14. **CAI Bibi; QIU Zhenyu; WANG Zhiwei**.
--- Topic Modeling For NIPS Words.

- Summary of the report：
This report uses statistical topic modeling to analyze the NIPS words dataset and identify the main topics of the papers. The authors use the latent Dirichlet allocation (LDA) model to model the generative process of each word in the document and identify topics based on the weights of each word. Clustering and dimension reduction methods like K-means, MDS, and tSNE are used to understand the relationships between topics and papers and track the popularity of topics over time.

- The report has several strengths, including:

- The report has several strengths, including:
The presentation of Latent Dirichlet Allocation and Fitting the LDA Model is very clear with equations and figures.

- The weaknesses of the report:
The method uses a bag-of-words model, which is not precise enough to model language. The conclusion about the topic trend seems suspicious because, in common sense, the neural network topic should be hot in recent years since AlexNet in 2012 and cold in the 90s.

- Evaluation on Clarity and quality of writing (1-5): 5

- Evaluation on Technical Quality (1-5): 4
languange model considering word order could be considered. Conclusion may need double check.

• Overall rating: 4.5
• Confidence on your assessment (1-3): 2

G16. **Chris HC Nguyen and James M Shihua**.
--- Robust PCA for Moving Object Detection in Video.
- Summary of the report：
This project primarily focuses on the use of RPCA to separate the foreground and background in a video. The dataset used for this purpose consists of CCTV videos, and the results have indicated the efficiency of the proposed method in accurately detecting and extracting moving objects from the background.

- The report has several strengths, including:
The methodology is clear, with a well-defined presentation of the algorithm's input and output, along with pseudo-code. The entire process is easily comprehensible.

- The weaknesses of the report:
The strengths and weaknesses of the method are not discussed.

- Evaluation on Clarity and quality of writing (1-5): 5

- Evaluation on Technical Quality (1-5): 4
languange model considering word order could be considered. Conclusion may need double check.

• Overall rating: 4.5
• Confidence on your assessment (1-3): 2

G05. **Yingxue XU, Jiaxin ZHUANG, Fengtao ZHOU**.
--- Visualization and Dimensionality Reduction Techniques for US Crime Data.
- Summary of the report：
The primary objective of this project is to utilize PCA for the purpose of visualizing the distinct categories within the US Crime Data. The author has

employed various techniques such as SPCA and MDS to serve as comparison methods for comparison.

- The report has several strengths, including:
The author employed several techniques including PCA, SPCA, MDS, ISOMAP, and LLE to compare the outcomes, which offer ample perspective to comprehend the effectiveness of PCA. The poster is intricately designed with a coherent flow of logic.

- The weaknesses of the report:
The conclusion is not informative as the author suggested that crime rates are influenced by all factors, yet no concrete experiments were conducted to validate different importance of all factors.

- Evaluation on Clarity and quality of writing (1-5): 5

- Evaluation on Technical Quality (1-5): 3
More detialed experiments are needed to show more informative conclusion.

• Overall rating: 4
• Confidence on your assessment (1-3): 3

G08. **MA Ruochen; Jihong TANG; Yuyan RUAN; Zhi HUANG**.
--- Ancestry Prediction via Dimensionality Reduction Techniques on SNPs Data.

- Summary of the report：
The main objective of this project is to assess the performance of seven different dimensionality reduction methods using the SNPs dataset. The author's findings indicate that the random projection method outperforms the other six approaches.

- The report has several strengths, including:
The assessment is adequate. The writer contrasted the outcomes produced by PCA, MDS, t-SNE, ISOMAP, LLE, UMAP, robust PCA, and random projection. And they show the superiority using quantitative metric.

- The weaknesses of the report:
In the comparison result, the raionality of the split scheme is not addressed.

- Evaluation on Clarity and quality of writing (1-5): 4

- Evaluation on Technical Quality (1-5): 4
More detialed experiments are needed to show more informative conclusion.

• Overall rating: 4
• Confidence on your assessment (1-3): 2