**Individual Report**

**Member name:** [Anuranjan Dubey](mailto:adubey37@asu.edu)

**Evaluated by:** [Sangeeth Santhosh](mailto:ssantho9@asu.edu)

**Date:** 10/14/2023

**Tasks Assigned:**

* Literature review for time series based machine learning techniques applied in: “TadGAN: Time Series Anomaly Detection Using Generative Adversarial Networks”
* Review and Evaluate [Rahul Nayak](mailto:rrnayak@asu.edu)’s in depth report.
* Review and Evaluate [Rahul Nayak](mailto:rrnayak@asu.edu)’s individual progress report.

**Summary:**

* This week I delved into identifying the difference in time series anomaly detection and traditional machine learning approaches discussed before.
* The time-series aspect can be helpful in real time systems also discussed in week 4 as real-time phishing tweet detection [22].
* TadGAN is an approach for detecting anomalies in time series data using GANs and LSTM Recurrent Neural Networks.
* TadGAN addresses the challenge of anomaly detection in time series data by training GANs with cycle consistency loss.
* It computes anomaly scores by combining reconstruction errors and critic outputs.
* Deep learning-based methods, such as LSTM AutoEncoder and MAD-GAN, handle complex temporal correlations a quick comparison and limitation are listed:
  + Traditional statistical models like ARIMA and Holt-Winters are used but require strong assumptions.
  + Dimensionality reduction with PCA is limited to linear reconstruction.
  + Auto-Encoder (AE) and Variational Auto-Encoder (VAE) capture latent structures in time series data.
  + LSTM Recurrent Neural Networks capture temporal correlations and predict future steps.
  + Generative Adversarial Networks (GANs) have the capability to capture complex temporal correlations and thus are successful.
* TadGAN outperforms baseline methods in most cases (8), offering high accuracy and generalizability.
* The paper's contributions include introducing TadGAN, evaluating it on 11 datasets, and providing insights into GANs for anomaly detection.
* TadGAN's relevance for real-time Twitter data is highlighted, but specific details about the Twitter dataset used are not provided in the paper.
* The paper used the "Tweets" dataset from the Numenta Anomaly Benchmark (NAB) in their evaluation of TadGAN.
* The NAB (Numenta Anomaly Benchmark) is a benchmarking system for time series anomaly detection that includes various time series datasets from different application domains.
* For real-time Twitter data analysis, time series analysis is more suitable to capture temporal trends and detect sudden shifts in Twitter activity.

**Outcome:**

* Traditional machine learning approaches might be less effective; they do not handle temporal data, and might miss insights in Twitter data.
* In cases, ML and time series combined, sentiment analysis could be applied and the resulting sentiment scores could then be analyzed using time series methods to detect sentiment trends.

**References** *(with citation)*

[3] A. Geiger, D. Liu, S. Alnegheimish, A. Cuesta-Infante and K. Veeramachaneni, "TadGAN: Time Series Anomaly Detection Using Generative Adversarial Networks," 2020 IEEE International Conference on Big Data (Big Data), Atlanta, GA, USA, 2020, pp. 33-43, doi: 10.1109/BigData50022.2020.9378139.

[22] Seow Wooi Liew, Nor, Mohd Taufik Abdullah, Razali Yaakob, and Mohd Yunus Sharum, “An effective security alert mechanism for real-time phishing tweet detection on Twitter,” Comput. Secur., vol. 83, pp. 201–207, 2019, doi:https://doi.org/10.1016/J.COSE.2019.02.004.

**Evaluation of Report  
  
Evaluation by:** [Sangeeth Santhosh](mailto:ssantho9@asu.edu) **Date: 10/15/2023**

**Is the weekly member report complete with all the major result(s) of the paper(s)? If not, provide as many examples of the major result(s) missing in the written report as possible. (in bullet form). [within 100 words]**

Yes, the weekly report is complete with all the major results of the research paper - “TadGAN: Time Series Anomaly Detection Using Generative Adversarial Networks”, covered in detail. The major results of the report help to understand the machine learning models alone would not suffice for anomaly detection in Twitter and hence a combination of sentiment analysis, machine learning and time series analysis is required.

**Is each section of the guidelines sufficiently completed? If not, point out what is missing. [Normally within 40 words].**

Yes, each section of the guidelines is sufficiently detailed and hence complete.

**Is the quality of this version of the written report satisfactory? If not, then why not? [Normally within 40 words]**

Yes, the quality of the report is good as it goes in detail into the major tasks covered by the reporter this week as well as a summary of the research paper.

**Approved by:** [Krupaben Kothadia](mailto:kkothadi@asu.edu) **Date: 10/16/2023**