Project Title: Anomaly Detection with Normalizing Flows for Tabular Data

1. Summary and contributions. Briefly summarize the project.

It was noticed that on the images data normalizing flows often assign higher likelihood to the out-of-distribution (OOD) data rather than the data which model was trained on. The authors try to answer the question whether normalizing flows do suffer from this problem when trained on tabular data. For this purpose, the authors conduct extensive experiments, considering several options of feature preprocessing and flow architectures. It is demonstrated that flows can provide decent performance in OOD detection on tabular data.

2. Strengths. Describe all the strengths of the project in enough depth.

The authors qualitatively and succinctly describe the motivation behind the work and provide references to the literature that preceded it. It is worth noting that the authors of this article pose research questions and provide detailed answers. Multiple experiments have confirmed that, in contrast to the images data, flows should not fail in anomaly detection on tabular data.

3. Weaknesses. Explain all the limitations of this project in enough depth.

There is no information about COPOD, ECOD, PCA, CBLOF, IForest methods the results of which are present in the Table 1. Also, it seems to me, it was worthwhile to indicate the description of the datasets used and what anomalies they contain.
Talking about simple encodings it would be better to explain in more detail where v_j came from. There is not enough detail in the description of the author's specificity of the latent representations of the VAE that were used as encodings, and how the work of Rezende D. and Mohamed S. is related to them.

4. Correctness. Are the claims and method correct? Is the empirical methodology correct?

The methodology of the project is correct.

5. Clarity. Is the project report well written?

The report is well structured. The report contains many graphs and tables illustrating the experiments. The work is replete with references to relevant literature.
It will be better to add contributions of each team member in details.

6. Related work. Is it clearly discussed?

The discussion of related work is sufficient and it is clear how the project is connected to the relevant papers.

7. Reproducibility. Are there enough details to reproduce the major results of this work?

The Github repo contains clear instructions how to run python scripts. However, the report contains poor information about the exact architecture of the used networks. The report and project GitHub repo lacks of details of (hyper)parameters such as batch size, learning rate, weight decay, number of epochs. Also, adding checkponits and notebook with the pipeline could be useful.

8. Overall score. You should NOT assume that you were assigned a representative sample of projects. The "Overall Score" for each project should reflect your assessment of the project.

Choose your score by **deleting** all the other scores.
(2) A very good submission; deserves high grade, tending to maximal (A).

9. Confidence score.

Choose your confidence score by **deleting** all the other scores.
(4) You are **confident** in your assessment, but not absolutely certain. It is unlikely, but not impossible, that you did not understand some parts of the submission or that you are unfamiliar with some pieces of the topic.