Credit Card Fraud Detection using Data Science

# **Chandramouli Yalamanchili DSC680 - T302 Applied Data Science (2215-1) https://chandu85.github.io/DSC680-Site/**

**Proposal Document** - <https://github.com/chandu85/data-science/blob/main/Project%201%20-%20Creditcard%20Fraud%20detection/Documentation/Project%201_Proposal.pdf>

# Week 2 check-in

Any surprises from your domain from these data?

* I have not come across any challenges or surprises with respect to the domain. As mentioned earlier the data is all transformed and normalized, so I don’t have a chance to really review the data for trends or data classification based on different features. As mentioned in last week’s proposal, I have a plan to review some of the references mentioned to see how a ML model can be integrated into credit card transactions real time processing. That could be a real challenge in my opinion. I will spend some time in next couple of ways to see what is the best option, if any to integrate the machine learning model into mainframe transaction processing.

The dataset is what you thought it was?

* Yes, data was what I thought it would be, I do see very less percentage of fraudulent transactions in data compared to the genuine transactions.
* One other thing I have noticed newly from the data analysis was that the transaction amount has very wide range of values. I have decided to normalize the amount column as well to avoid the wide range it currently has.
* The might run some data analysis on time and amount fields as well into next week to see if there is any correlation between transaction time, amount and the fraudulent transactions.

Have you had to adjust your approach or research questions?

* Not really, at least at this time I am still going ahead as per my plan at this time.
* I had to do some additional code to normalize the amount field, which was very quick.
* I might have to do something into next week while model training to ensure I take care of the skewness I have in the data (very less number of fraudulent transactions)
* Also as per the reference I was using, I have seen that they have compared models build with different methods, that looks interesting to me. Earlier I was just going to build one model, I might go for building different models and compare their performance.

Is your method working?

* Yes, I would say my initial plan is still working at this time.
* As mentioned in previous response, I might go ahead and do a little bit more than initially planned, by building multiple models rather than just building one like initially planned.
* I have planned to use Python notebook and haven’t had any issues with it so far. Everything has been going as per the initial plan.

What challenges are you having?

* I haven’t run into any challenges yet either with technology or with data. Everything has been going pretty smooth so far for this project.
* I might have to spend some additional time into next week based on my analysis for Time and Amount fields and if they give me some insights that indicates the need for me to use those fields as-is, for example, if I see more fraudulent transactions during night-time, I don’t want to exclude time as a feature in the model. Same for amount, if I see more fraudulent transactions at lower dollar amounts, I might have to see using amount data as-is might be of more value.
* So, at this time everything is going smooth for me, but next week might be interesting.