

Dear Candidate,

Please find all instructions needed to complete your assignment below

**General guidelines**

1. All coding must be done in python 3 in a jupyter notebook.
2. All code from external sources should be cited.
3. Please explain your thought process clearly using comments.
4. Final Submission will be a jupyter notebook in .ipynb format and .html format.
5. You may use any tools, github repos, python libraries etc but please make sure to cite them clearly.
6. You may use any computational resource for completing these tasks.
7. Please ensure that your notebook runs before submitting. Also indicate the conda and python package versions used as a requirement.txt file or within the notebook itself.
8. You may use free tier google cloud, azure or colab notebooks for doing this assignment. Kaggle also offers free gpu.
9. The datasets used for the assignment are open source and plenty of work on them could be found online. We basically would like to see how you approach the tasks below.
10. A single Zip file submission with all notebooks, models should be sent to the contact below. You may use the same email below for any clarifications.

**Dhanush Dharmaretnam,  
Lead Data Scientist, Sgsco Labs  
dhanush.dharmaretnam@sgsco.com**

**What we are looking for ?**

1. Simplicity and innovation.
2. Ability to deploy ML models to cloud
3. Building Scalable Solutions
4. Ability to debug and solve bugs.
5. Coding skill.
6. Follow coding standards (pep-8).

## **Datasets**

You are provided with

- A) Credit card fraud detection dataset (Classification Task)

## **Task**

The credit card fraud detection dataset [Standard kaggle dataset]. You may reference existing kaggle notebooks. We are looking for how you approach this problem rather than achieving top accuracy.

<https://www.kaggle.com/mlg-ulb/creditcardfraud?select=creditcard.csv>

Feature 'Class' is the response variable and it takes value 1 in case of fraud and 0 otherwise. More description of the dataset could be found in the kaggle link above.

## **Task description**

- 1) **Train a classifier to predict if a transaction is a fraudulent transaction or not ?**
- 2) **How would you deploy this model in production? (Explain the approach)**
  - a) **Deploy the model as an api endpoint to any cloud provider. Most of them offer free credits when you sign up.**
  - b) **API response should be clear and should handle missing data and exceptions.**
  - c) **You are expected to share how your model could be called.**

**Example of a sample api call below:**

```
import requests
import time
import json
data = {"feature1":1.0,"feature2":2.0,"feature3":3.0,"category1":"Toronto",
        'calltype':'test'}
start_time = time.time()
r = requests.post("www.mypredictor.com/predict/", json = data)
print(time.time()-start_time)
print(r.json())
```

**The above is an example only. You may deploy to kubernetes, function/serverless apps or flask on a VM. A successful app will offer insights in terms of logs and load. It should be able to scale**

