**Decode 11.1**

**Round – Data Modelling and Machine Learning**

Problem Statement

**Data Modelling** is the creation of models on the basis of past and present data to predict the future instances. A data model trained perfectly is able to predict the results accurately. During the creation of a model the main steps involved are: data gathering, feature engineering, features selection, modelling, parameter tuning and metrics evaluation. Each of the steps are important in data modelling.

**Machine Learning** plays a vital role in training of the data. Machine Learning algorithms are powerful enough to create a best classifier. The selection of a particular algorithm depends on the data and features which are in use to train the model. Although there is no fixed method to choose a particular model.

In this work, you have to create a best classification which is able to predict the income a person. The data required for this is attached in the data (train.csv). We expect you to utilize the current large volume of research and development in this area to analyse the given use case of classifying personal income. The complete description of dataset variables can be found in the **data\_desc.txt** file provided with the data. We hope that you will come with a good model in terms of accuracy to classify the income.

Following work must be present:

* A model with high robustness and accuracy.
* Definite procedure for creating the model
* Visualisations to reflect data and feature importance
* Visualisation to estimate model

**You can use any model/library.**

Tech Stack Recommendations:

Analysis:

* Excel
* Seaborn/Matplotlib

Development & Processing:

* Sklearn
* Pandas
* Numpy
* Tensorflow/Keras (Optional)

Resources:

* [Seaborn](https://seaborn.pydata.org/)
* [Sklearn](https://scikit-learn.org/0.21/documentation.html)
* [Matplotlib](https://matplotlib.org/stable/contents.html)
* [Numpy](https://numpy.org/doc/)
* [Pandas](https://pandas.pydata.org/docs/)

**Note:**

* You aren’t restricted to any tech-stack, all the above info is just to help you get started. You are encouraged to go creative with your implementation and use whatever technologies you find interesting or want to try out.
* Don’t forget to spare some time for preparing your presentations before the end of the hackathon.
* All hackathons target only a prototype proof-of-concept level development.

**Keeping that in mind we wish you best of luck and expect great outputs from each and every team.**