

**Group Members** : Sergii Deshunin, Gwanghyun (Gina) Jung, David Chong

**Data Set** : <https://www.kaggle.com/ellipticco/elliptic-data-set>

**Project Aim** : Develop models to classify licit and illicit transactions in bitcoin network

### **Why is it interesting**

As per the associated paper's background, being able to identify licit and illicit transactions will enable more robust regulation of the use of the bitcoin network and possibly other digital currency networks. This will help to boost confidence in the legitimacy of these currencies and also improve equitable access to these financial services.

Additionally, the use of graphical data is a relatively untapped area in machine learning. Learning how to represent graphical data in more traditional forms and also how new network structures like Graph Convolutional Networks function will be very educational in its own right. Furthermore, the set of unlabeled data provides the opportunity to learn about semi-supervised learning as well.

### **Challenges**

The dataset is imbalanced (10% Illicit cases vs. licit) and still around 75% of data are unlabeled. Some of the features in the dataset are anonymised, which will make it harder to understand the data. Additionally, GCNs and other network structures for graphical data are probably newer and less accessible in terms of understanding and package implementations than other models.

### **Types of models**

Logistic regression

SVM

Random forest

GCNs and other graphical neural networks

### **Roles**

Gwanghyun Jung - EDA

Sergii Deshunin - Explore non-deep learning models

David Chong - Explore deep learning models