ML - Assignment :
Machine Learning in Blockchain Systems

Name : Ashwin Vaswani
Institute: Birla Institute of Technology and Science (BITS Pilani) , Goa Campus
Email : ashwin.vaswani99@gmail.com / f20170960@goa.bits-pilani.ac.in
Contact : +91 7977572467

Problem Statement :

"A blockchain is a time-stamped series of immutable record of data that is managed by a cluster of computers not owned by any single entity. Each of these blocks of data (i.e. block) are secured and bound to each other using cryptographic principles (i.e. chain)".

Architecture : These are the core blockchain architecture components:
1. Node — user or computer within the blockchain
2. Transaction — smallest building block of a blockchain system
3. Block — a data structure used for keeping a set of transactions which is distributed to all nodes in the network
4. Chain — a sequence of blocks in a specific order
5. Miners — specific nodes which perform the block verification process
6. Consensus— a set of rules and arrangements to carry out blockchain operations

Blockchain possesses a lot of benefits for businesses. Here are several embedded characteristics: cryptography, immutability, provenance, decentralization, anonymity, transparency. A layered approach of understanding may be easier for the task.

Steps in a blockchain application: The actors such as users and IoT devices interact with a typical blockchain application, which in turn interacts with a blockchain network. Before this flow starts, the blockchain network and governance, including peers, membership services, and endorsement policies, have already been designed, implemented, deployed and in Operation.

Task :

1. Select one or more components in the architecture of the blockchain system where you can apply any machine (deep) learning technique(s). The task is to use machine learning to enhance/break the fundamental blockchain system.
2. Demonstrate your idea in the following ways (both are required):
1. Explain it

i.What is the problem statement?
ii.What is the proposed solution?
2. Implement a prototype to show “Yes, it works!”
i.Mine data from the web
ii.Perform exploratory data analysis
iii.Extract features
iv.Train a few models, perform cross-validation/hyper-parameter tuning
v.Report the robustness of solution by improvised metric scores
vi.Compare with previous models/metrics (if any)
3. Some of the details are left open-ended for you to figure out and use your strengths. Please don’t use blockchain to enhance machine learning applications. The task is to use machine learning to enhance/break the fundamental blockchain system.

Overview:

The problem statement required me to use a machine learning / deep learning model to enhance the fundamental blockchain system. I was able to use Machine Learning on an architectural component that is the Nodes in a Blockchain to enhance the transactions in a Bitcoin OTC Blockchain network. The solution proposed deals with the Bitcoin OTC data to predict the rating a source node gives a target node in a network. As mentioned, the architectural component of blockchain on which machine learning techniques have been used here is the node. Out of all the models I implemented, I achieved a maximum accuracy of 60.5% on 21 way classification task to predict ratings between -10(Most Distrust) to +10(Most Trust)

How to run :

I have provided a jupyter notebook in the zip file attached which contains all the necessary code.

Methodology:

The first and major part of this assignment was to understand the architecture of blockchain and how it works so as to think of ways in which machine learning can help to enhance Blockchain. I read multiple papers and had multiple ideas in mind such as illegitimate transaction detection and bitcoin transaction price detection etc. I then read about Bitcoin OTC (Over the counter) and understood the mechanism. Bitcoin OTC consists of users in a network that do transactions with each other and trust is a major factor. I decided to use Machine Learning to predict the

rating a source node gives to a user node. I browsed the web and searched for relevant data that can help me solve this problem. I used the Bitcoin OTC dataset made available by Stanford University for this problem. Link : https://snap.stanford.edu/data/soc-sign-bitcoin-otc.html

The dataset contains 4 columns namely source , target , rating and timestamp. I decided to convert this into a graph based problem to do some feature engineering to extract features. I used NetworkX to handle the graph created with users as nodes or vertices and ratings between them as the edges. With the help of this graph, I created two features namely **Fairness** and **Goodness.** Basically, if a user gives similar ratings to all other users that he rates, it can mean that the user is not fair in his ratings. Goodness on the other hand measures the quality of a user while being rated. I used these two features, the fairness of a source user and the goodness of the target user to make predictions for the rating a source user may give to a target user in a Bitcoin OTC network when no prior transaction has been made. This will help the user to understand if he should or should not do a transaction with some particular user and thus improving the security of blockchain transactions in a Bitcoin OTC network.

Here is a table with all the new results that I got using different Machine Learning models:

Model / Accuracy	Validation	Test
Logistic Regression	58.8%	59.3%
Random Forest	54.1%	54.6%
Gradient Boosting	59.6%	60.5%

Conclusion:

As we can see from the above table, the gradient boosting model which is like an in-place ensembling technique achieves the best accuracy in predicting ratings between -10 and 10 which is a 21 way classification task on the ratings a source user may give to a target user. Since in a Bitcoin OTC Network, trust or ratings is a major factor , these predicted ratings can be very helpful in helping a source user visualise compatibility with a user he hasn't transacted with yet. Due to time and resource constraints, machine learning techniques were used in this report. However, I am sure that the model can be further optimised by trying Neural Networks and Graph Neural Networks since data is so vast. In conclusion, through this assignment we are able to use Machine Learning techniques on an architectural component of Blockchain (Nodes in this case) to enhance Blockchain transactions on Bitcoin OTC Network.

Libraries / References used:

I used sklearn's logistic regression initially and then moved onto ensemble techniques like Gradient boosting,Random Forest etc. from sklearn.ensemble. I used numpy and pandas for basic data handling operations and seaborn and matplotlib for visualisations. I used NetworkX to handle graphs.
Here is a list of some of the papers I read during the time period of this assignment :

https://cs.stanford.edu/~srijan/pubs/wsn-icdm16.pdf
https://pdfs.semanticscholar.org/2ea6/04d967ca11ec869545ace248c41db6a49855.pdf
https://journals.plos.org/plosone/article/file?id=10.1371/journal.pone.0207000&type=printable
http://snap.stanford.edu/class/cs224w-2014/projects2014/cs224w-20-final.pdf