ABSTRACT

With the rapid development of the internet and the amount of content circulating on the internet, IP networks are considered less reliable in handling content-based data exchange. Therefore, a new network architecture called Named Data Networking (NDN) was created which changes the host-centric communication model to be data-centric. NDN has a storage area that serves to store a cached copy of the data. The data-centric NDN structure makes the nodes on the NDN have a high load and can reduce NDN performance.

This final project research performs NDN-based dataset retrieval, pre-processing the dataset and applying machine learning to predict the load on network nodes. By using the network data from the ndnSIM simulation on the GEANT topology and using the Machine Learning Item Average, User Average, Global Average and User-Item Average models provided by Orange. Load prediction is carried out with 3 dataset scenarios, 3 feature selection scenarios and 4 types of Machine Learning models for each dataset.

The results of the NDN network dataset retrieval amounted to 1,370,512 rows of data containing NDN network traffic. The NDN network dataset is reduced by 83% after the pre-processing stage. The results of the load prediction on the NDN node using Machine Learning show that the best prediction is obtained when using the Item Average model with the selection of the Node-Type and dstNode-Type features, resulting in an MSE of 0.013 for all scenarios.

Keyword: NDN, Machine Learning, weight on node, Orange