# Network traffic analysis using machine learning

## **Abstract**

Recent development in smart devices has led us to an explosion in data generation and heterogeneity, which requires new network solutions for better analyzing and understanding traffic. These solutions should be intelligent and scalable in order to handle the huge amount of data automatically. With the progress of high-performance computing (HPC), it becomes feasible easily to deploy machine learning (ML) to solve complex problems and its efficiency has been validated in several domains (e.g., healthcare or computer vision). At the same time, network slicing (NS) has drawn significant attention from both industry and academia as it is essential to address the diversity of service requirements. Therefore, the adoption of ML within NS management is an interesting issue. In this paper, we have focused on analyzing network data with the objective of defining network slices according to traffic flow behaviors. For dimensionality reduction, the feature selection has been applied to select the most relevant features (15 out of 87 features) from a real dataset of more than 3 million instances. Then, a K-means clustering is applied to better understand and distinguish behaviors of traffic. The results demonstrated a good correlation among instances in the same cluster generated by the unsupervised learning. This solution can be further integrated in a real environment using network function virtualization.