A Study of Supervised Machine Learning Algorithms for Traffic Prediction in SD-WAN

Authors Kashinath Basu, Muhammad Younas, Shaofu Peng

Abstract

Modern cloud, web and other emerging distributed services have complex network requirements that cannot be fulfilled via classical networks. This paper presents a novel architecture of a noble Software-Defined Wide Area Network (SD-WAN) that provides the framework for incorporating AI/ML based components for managing different centralised services of the WAN.To leverage the benefit of this framework, a crucial early stage requirement is to accurately identify the traffic category of a flow based on which follow-up actions such as QoS provisioning, resource orchestration, etc. can be implemented. To address this, the research then presents the model of a supervised ML based traffic prediction module and presents a detailed comparison and performance analysis of a shortlisted set of ML models with a variety of traffic categories. The research also takes into account the serialized processes in the models' training and learning phases emphasizing on the sensitivity of the feature selection process in the performance of these algorithms.

Keywords—supervised machine learning, ML, artificial intelligence, AI, software defined network, SDN, SD-WAN, QoS, QoE, feature selection, naïve bayes, decision tree, nearest neighbor, support vector machine, logical regression

1 Introduction

There has been a continuous evolution of networked applications of different categories with diverse range of QoS requirements. This has further accelerated with the increase in the number and variety of cloud and web services over the Internet. Some of these application categories include immersive metaverse video and audio, VoIP, email, social media, peer-to-peer (P2P) data transfer, web browsing and search, inter and intra cloud synchronisation traffic and so on. The QoS requirements of these application categories vary significantly. For example, an immersive video application will generate video frames periodically at a predictable rate and will require a low delay, low jitter and low loss guarantee from the intermediate network to provide the appropriate level of quality of experience (QoE) to the end users. In contrast, a web browsing application will generate bursty traffic and be able to tolerate longer end-to-end delay and can recover from loss using retransmission. In order to provide the appropriate QoS to the different categories of applications with optimal resource utilisation, the intermediate network has to be traffic engineered such that it can provide customised QoS based on the traffic category.

The intermediate network on an end-to-end path of an application is generally heterogeneous and can be broadly divided into three main tiers of network: access, distribution and the core wide area network (WAN). A tier may be further subdivided based on the complexity of the network at that level. Generally, access and distribution tiers lie within the boundary of an enterprise or may extend up to the local Internet provider. Here, the number of aggregated sessions are restricted to only those starting or ending in those networks, the QoS requirements and the traffic profile of the applications are mostly known in advance and network resources are generally cheaper and available in abundance. Because of these advantages, QoS provisioning is comparatively simpler in these two tiers of network.

In contrast, the core tier is significantly complex. It is composed of one or more intermediate WANs. A WAN will have several times more traffic joining from a large number of distribution tier networks wherein aggregated traffic will come from a more wider pool of applications with a more heterogeneous set of QoS requirements. Since the major segment of the end-to-end journey of the application traffic is through the WAN, therefore it is critical that the transit traffic through the WAN receives the required level of QoS to provide the desired QoE experience to the end users.

A WAN is composed of a large number of forwarding devices of different processing capabilities interconnected by links running at different speeds. In this setting, providing adequate QoS to an application session involves in-depth understanding of the QoS requirements of the application at the ingress point of the WAN and then immediately orchestrating and provisioning these resources throughout the network. These steps are however difficult to implement in a traditional legacy WAN due to a number of bottlenecks. Firstly, to add capacity or to partition existing resource pool for an application will involve at the minimum manual reconfiguration of the networking devices and repartitioning of link bandwidth and other resources. These are time and cost intensive processes that need expertise and could be error prone if not done with adequate planning. Secondly, the distributed nature of operation of the legacy WANs restricts a WAN wide view of the state of the topology at any point. Therefore, among the large number of links and devices that constitute the WAN infrastructure, when some of them occasionally fails, the current signalling mechanism in the WAN routing protocols are slow in propagating these state updates to the edges. Hence, any route planned at the edges could be based on an outdated topology. Thirdly, legacy WAN routing protocols such as Routing Information Protocol (RIP) and Open Shortest Path First (OSPF) (Ghosh et al., 2001) do not give significant weightage to QoS metrics and therefore are inadequate to provide the QoS guarantees based on the application's requirements. The challenges are further aggravated by the fact that the current approaches for interpreting the QoS requirements of the applications at the edges of the WAN are themselves inadequate to accurately identify the different categories of applications and their precise QoS requirements (Xue et al., 2013). Due to these issues, WAN operators adopt a simplistic but expensive and wasteful approach of over provisioning the resource pool in the network to ensure that all traffic irrespective of its QoS requirements receives a low delay, low jitter and low loss service. This results in typical WAN links and devices running at only 30-40% utilisation even during peak period wasting significant amount of resources (Bogle et al., 2019).

However, the emerging trend of a softwarised networking infrastructure in SD-WAN can be adapted to address these challenges and many others by facilitating a centrally managed programmable framework for hosting a range of services starting with the prediction of the ingress application category. The key novelty of this research is the extension of this programmable WAN framework by taking advantage of the recent developments in the field of supervised ML algorithms and then its application in the domain of network traffic pattern analysis and the prediction of the application category. The main contribution of this research can be summarised as follows:

- A design of an SD-WAN framework focussing on the traffic prediction aspects.
- Review and analysis of current research in the field of traffic classification and prediction.
- A study of the kaggle dataset used in this research from the perspective of feature selection for the traffic category prediction by the ML models.
- Analysis of performance metrics for comparing performances of ML models.
- A thorough analysis of prediction performances of a shortlisted set of ML models based on the metrics identified.

The reminder of the paper is organised as follows: At first, in section 2, we present the review and analysis of the existing research in the field of network traffic prediction and its pros and cons, focusing primarily on the application of ML for the prediction process. In section 3 we present the overview of the proposed SD-WAN framework that incorporates the traffic prediction module. We discuss the dataset and the feature selection process used in this research in section 4. Selection of appropriate performance metrics is crucial for accurate comparison of the ML models. In section 5 we present an analysis of the performance metrics used in this research. Section 6 presents an analysis of the results and evaluation of the performance of the ML models for predicting the traffic category of the network traces; finally, section 7 concludes the papers and summarises key features and novelty of the work.

2 Related work

This section reviews existing work related to network traffic classification. Traffic classification techniques can be broadly divided into three categories. These include port-based classification, classification based on packet inspection and flow and session based classification. The port-based classification scheme is the simplest and oldest technique for identifying traffic category based on the transport layer port numbers and protocol type. In theory, if there were only a limited range of applications confirming to the Internet Corporation for Assigned Names and Numbers (ICANN) well-known and registered address allocation scheme (IANA 2023), the port-based scheme would have been sufficient. However, in reality the range of applications on the Internet have grown significantly resulting in dynamic port allocation as well as applications ignoring the port-numbering scheme. This makes port-based traffic classification difficult (Moore and Papagiannaki, 2005). Furthermore, with IPv6 and virtual extensible local area network (VXLAN) tunnelling (Mahalingam et al., 2014) in modern enterprise traffic, the transport layer addressing have become even less credible to identify the application sources of the traffic. However, port-based classification could be used in conjunction with other more robust schemes for cross validation purpose.

In classification based on packet inspection, the classification takes place based on the analysis of the content of the payload. Though this process provides good level of accuracy (Bremler-Barr et al., 2014), it is time and resource intensive. In addition, given that a significant proportion of modern network traffic are encrypted for security purpose, the deep packet inspection (DPI) technique would become more difficult. There are schemes that try to handle this issue by inspecting only the initial unencrypted part of Transmission Control Protocol (TCP) connection setup before the actual transmission of application level data (Wang 2015). However, their accuracy is inconsistent and they are also unable to classify User Datagram Protocol (UDP) flows which constitute a significant part of streaming and multimedia traffic in the Internet. Another practical challenge is the fact these schemes require priori knowledge of the application classes so that traffic can be tagged to these classes based on their behaviour. This requires packet based classification schemes to always maintain an updated database of all existing and emerging application classes and their behaviour. However, this category of classification is suitable in certain scenario where time and resource are not constraints; for example, in offline classification for retrieval of ground truth information (Gringoli et al., 2009) or in cases where it is used infrequently such as being the last resort in a multi-tier hierarchical classification scheme (Schueller et al., 2018).

Flow and session based traffic classifiers use behavioural and statistical traffic features such as packet length and interarrival time distribution, duration of the communication, amount of data transfer, etc. and other derived values such as mean and standard deviation of the features to predict the traffic category. These features could be captured at the level of a unidirectional flow between a source-destination pair, or a session which is a pair of unidirectional flows in opposite direction. A combination of five tuple pair from layer 3 and layer 4 headers, that comprise source and destination IP and transport layer addresses along with transport protocol, is used to identify a flow. In IPv6 networks, the flow label field can also be used instead for this purpose if the IPv6 packets are marked accordingly. A big advantage of this category of classification is the fact that it does not inspect the content of the application layer payload and can therefore operate transparently even over encrypted traffic. In addition, it can get around the unreliability of transport layer dynamic ports numbers used in port-based classification. The temporal and spatial features of the traffic can be readily captured from any suitable point in the network and analysed thereafter. There are a range of supervised and unsupervised ML algorithms under this classification scheme which provide different levels of accuracy and are based on analysis of different attributes of the source (Santiago Lopes Pereira et al., 2014; Liu *et. al*., 2020; Jose, 2022).

Most of the above research attempts to identify individual applications rather than the general application category. With new applications evolving regularly, the database of the application may not always be up-to-date. This approach is therefore not scalable across the Internet within a short duration of time. This approach is more challenging and irrelevant from the QoS provisioning point of view where the general application category is more important for identifying the resource requirements and orchestrating the traffic engineering aspects. In earlier research (Basu et al., 2002) it has been seen that applications of the same category have similar traffic characteristics and require identical QoS. The traffic flows can therefore be grouped together under the same category label and service at the coarse granularity of the traffic class rather than the fine granularity of individual flows thereby making resource orchestration and provisioning scalable. This has been an important feature of differentiated services and later multiprotocol label switching (MPLS) (Akyildiz et al., 2003) and has been proposed as the basis for traffic engineering in our SD-WAN framework. However, as the first step towards this approach, it is important to accurately identify the application category of the traffic sessions entering the network so that they can be grouped together with similar flows. Incorrect classification will result in flows grouped with wrong traffic categories and thereby not receiving the right level of QoS. For example, if a video session is classified as a file transfer session and grouped accordingly, then video will not receive adequate realtime QoS bandwidth and delay guarantees resulting in unacceptable delay, jitter and loss, thereby degrading the end-to-end application performance resulting in poor user experience. Table 1 shows a summary of the traffic classification techniques.

TABLE 1 SUMMARY OF TRAFFIC CLASSIFICATION TECHNIQUES

Classification Category	Features	Research Ref	Strength	Weaknesses
Port based	• Uses the ICANN assigned port numbers.	• Mahalingam et al., 2014. • Moore and Papagiannaki, 2005. • Azab et al., 2022	• Easy to retrieve and simple process	• At times, applications ignore the ICANN port numbering scheme. • IPv6 and VXLAN tunnelling makes port addresses even less credible
Classification based on packet inspection	• Based on deep application layer payload inspection • For encrypted traffic, the initial TCP exchange	• Bremler-Barr et al., 2014 • Wang 2015 • Gringoli et al., 2009 • Schueller et al., 2018 • Bujlow et al., 2015 • Ghosh and Senthilrajan, 2019	• Good level of accuracy • Suitable for ground truth information • Last resort in a multi-tier hierarchical classification	• Time and resource intensive process • Encrypted traffic cannot be deeply processed
Flow based classification	• Uses temporal and spatial features of the flow such as packet length , packet interarrival time and flow duration distribution	• Santiago Lopes Pereira et al., 2014 • Liu et. al., 2020 • Basu et al., 2002 • Khater et al., 2015 • Salman et al., 2020 • Jose, 2022	• Transparent to port numbering and traffic encryption	• Feature may change as traffic propagates through the network

In the context of a SDN-WAN, resource orchestration and reservation are centralised and carried out on a network wide basis based on the classification of each of the sessions at the ingress points of the network (Yang et al., 2019). An incorrect classification in this context have a more wide scale network wide repercussion unlike in a traditional network where QoS resource allocation is more locally restricted to the individual routers or small subnets (Basu et al., 2020).

In order to improve the accuracy of network traffic classification, in recent years, ML models have been proposed for predicting the application category and sometimes even the specific application names with varying degree of success (Boutaba et al., 2018). In the context of this research, predicting the individual applications do not bring any additional benefit. As mentioned earlier that the core and distribution level network only needs to know the application categories so that identical applications can be grouped and serviced together based on their QoS requirements. The focus here is on improving the accuracy of the prediction process of the application category levering on the strengths of the different ML models.

In the next section, we first present the design of the SD-WAN model that provides a framework for building different components for various functionalities and operations for optimising and enhancing the performance of the SD-WAN. Within this model, we then focus in detail on the design of a traffic prediction module to demonstrate how supervised machine learning can be used for traffic prediction by leveraging on the architecture of this noble framework.

3 THE PROPOSED SD-WAN FRAMEWORK

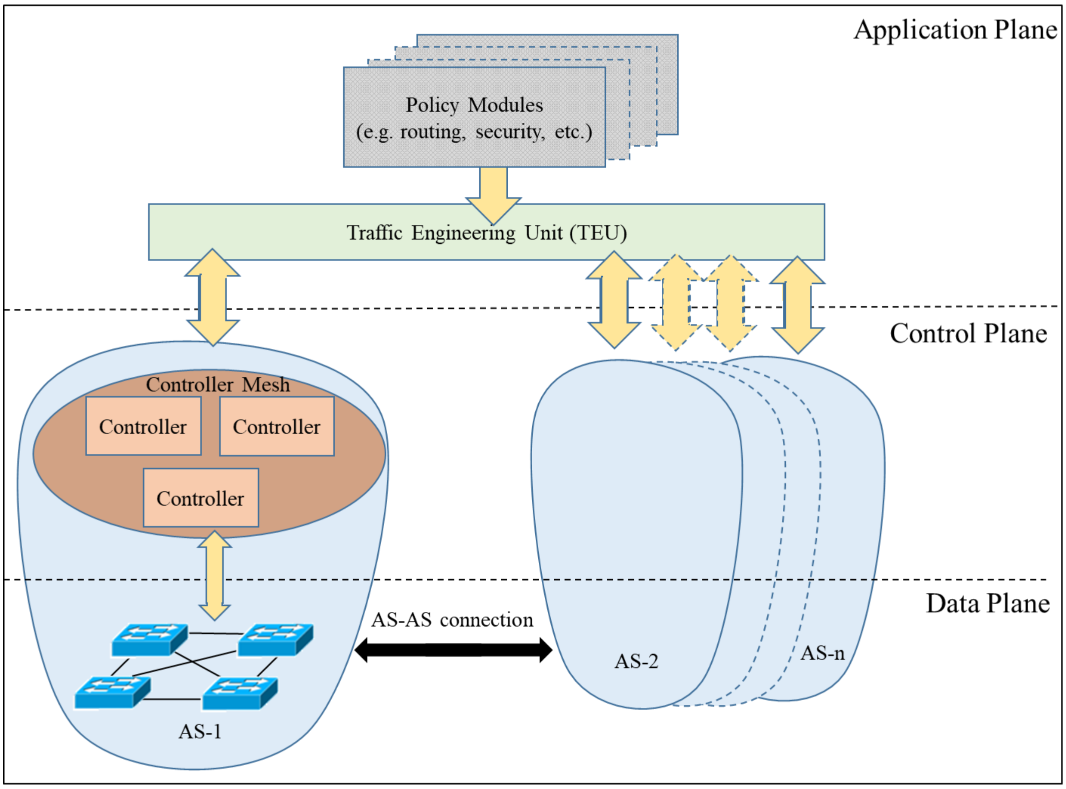


Figure 1 Architecture of the SD-WAN framework

Our proposed SD-WAN framework extends the general SD-WAN model and is composed of three planes (Fig. 1). At the bottom layer is the data plane that is primarily made of a mesh of high-speed merchant switches. This layer focuses primarily on the low-level optimisation of the switch forwarding pipeline and its associated forwarding tables in order to provide a high-speed switching fabric across the WAN. The actual switching decisions are however made at the layer above in the control plane that is made up of one or more controllers and is connected to all the switches in the WAN. The control plane supplies the high level switching policies to the data plane via structured southbound signalling primitives using well-defined signalling protocols such as Openflow (Hu et al., 2014) or Forwarding and Control Element Separation (ForCES) (Latif et al., 2020).The switches in the data plane configures their forwarding tables based on these information. The decision making in the control plane is carried out based on two types of input. The first type of input comprises information about the current state of the network such as link or device outage and utilisation levels — these are supplied directly by the data plane in the south via the same signalling mechanism. The control plane receives the second type of input from the north from the application plane that is comprised of specialised applications providing specific services to the SD-WAN and running on top of a traffic engineering overlay service (Fig. 1). In large SD-WAN for scalability purpose, the network can be comprised of separate autonomous systems (AS) with its own data and control plane and its associated switches and managing controllers. On top of this, there is a single overlay network wide application plane with the centralised network wide view and running network wide applications. Our focus in this research is on a proposed ML-based Traffic Prediction Module (TPM) that will operate as part of the Traffic Engineering Unit (TEU) in the application plane as shown in figure 2. The TEU receives input from the policy modules viz. security policy module and routing policy module, SLA policy module, etc. and orchestrates and provisions QoS to the labelled aggregated traffic.

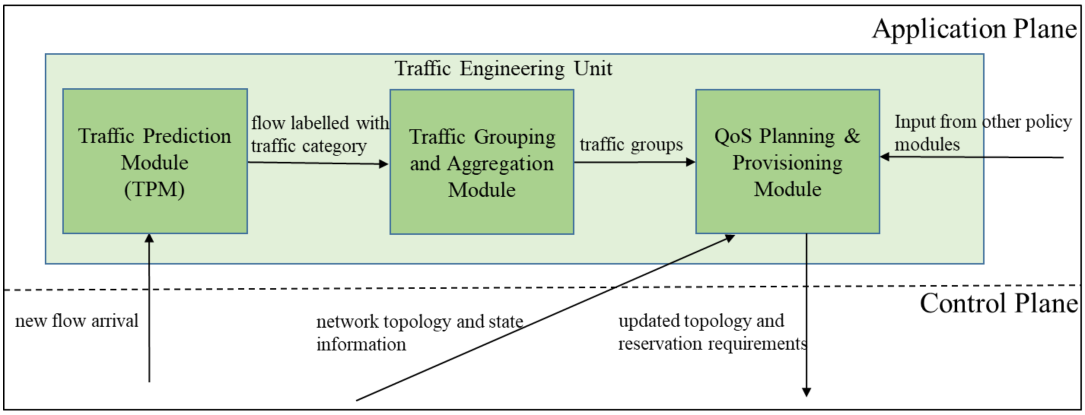

Application Plane
Traffic Engineering Unit
Traffic Prediction Module (TPM)
flow labelled with traffic category
Traffic Grouping and Aggregation Module
traffic groups
QoS Planning & Provisioning Module
Input from other policy modules
Control Plane
new flow arrival
network topology and state information
updated topology and reservation requirements

Figure 2 Components and interactions of the Traffic Engineering Unit with the rest of the SD-WAN framework

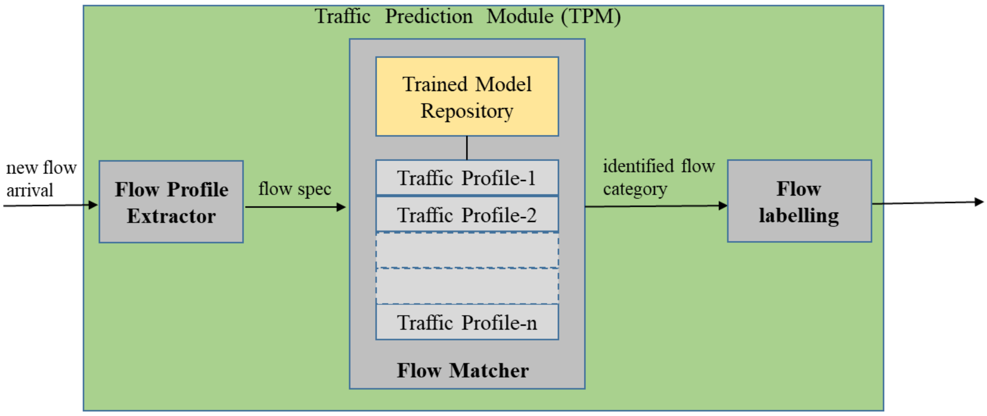


Figure 3 Different sections of the Traffic Prediction Module (TPM)

The role of the TPM is to inspect flows arriving at the ingress points of the SD-WAN and to classify them into one of the predefined set of traffic categories (Fig. 3). The classified flows are then labelled and send to the next module for grouping and aggregation. For the classification purpose, the TPM uses a supervised ML approach. Inside the TPM, there is a repository of a range of traffic profiles and their associated features. A suitable ML model trained with a shortlisted set of apt features from the traffic profiles is then used for prediction of the application category of the input flow. In this research we attempt to identify the combination of the minimum number of apt features and a corresponding suitable trained ML model that will provide the highest level of accuracy in the prediction process. The objective, to identify only the minimum set of relevant features, is that we can optimise the time required both during the learning phase and later on during the prediction phase in the live system. Later, in section 6 we present a detailed analysis of the performance of a shortlisted set of supervised ML models for its suitability in the prediction module. Note that the objective of the classification by the TPM is to identify traffic categories and not individual applications. There are several advantages of this approach in terms of its practicality and scalability. From the network's perspective, QoS has to be provisioned based on the characteristics and requirements of the traffic and the knowledge of the application is not directly relevant. There are wide range of applications that require identical QoS. By grouping them based on their traffic category and servicing all the flows within that category together rather than individually is a more scalable solution from the network resource orchestration and reservation perspective. Moreover, there are continuously new types of applications joining the global Internet infrastructure. It is therefore difficult to keep track and train the ML model on a real time basis to keep it up-to-date with these new applications. In addition, for the prediction process, it is more challenging and error prone to predict individual applications rather than application category, especially when applications have similar traffic characteristic. For these reasons, our TPM focuses on classifying incoming flows into a number of predefined categories. In this research, we have focussed on a number of broad traffic categories by grouping related types of applications that share identical QoS. These categories and the corresponding applications are presented in table 2. The TPM can be extended to include new categories without making any changes to the overall SD-WAN framework.

The output from the prediction module will be labelled flows which can then be grouped with other identical flows that belong to the same traffic category and the aggregated QoS requirements for each category is then updated accordingly (Fig. 2). The grouping can be more fine-grained and may also include additional metrics such as source and destination addresses such that flows of the same category and belonging to the same route only are grouped together for further operation. The traffic groups and their associated QoS requirements are then passed to the next module for QoS planning and provisioning. This QoS planning module will also take input on the current network state and topology as well as any specific configured policies such as QOS based routing policy, security policy, etc; this is to compute the updated routing and reservation requirements. These could be computed individually for each AS of the SD-WAN at a high-level of abstraction and then passed to the corresponding control plane of each AS. At the control plane, the QoS requirements are mapped to individual switch level configuration requirements and are passed to the data plane switches using southbound signalling. The switches will then use this information in order to update their forwarding tables and to implement the desired changes in the topology. This overall process is based on processing and mapping of control information across the three planes with traffic prediction acting as the first trigger in the multi-step process. Hence, the reliability and accuracy of the prediction process in the TPM are crucial for the correct operation of the follow-on modules in the application plane and for meaningful mapping of the control information in the control and data plane of the proposed SD-WAN framework. The design of the TPM is centred on a suitable ML model trained with a range of traffic categories with a shortlisted set of optimal features. In the next section, we present an analysis of the traffic categories and their features used in the dataset for training the shortlisted ML models.

4 DATASET AND FEATURE SELECTION

The heterogeneity of the dataset, the accuracy of the ground truth and the features available in the dataset are important for shortlisting relevant features to facilitate good quality learning of the supervised ML models.

4.1 Dataset

The traffic trace was sourced from the kaggle repository (Rojas et al., 2020; Rojas 2017) and have over 3.5 million sessions from seventy five different applications. This heterogeneity of the dataset and the large number of session traces available can help in reducing the overfitting and underfitting problems of the models during the learning phase. The application included iCloud, iTunes, Dropbox , Facebook, Gmail, Google suite, Google Maps, HTTP, Instagram, Microsoft One Drive, Facebook Messenger (MSN), Netflix, Skype, Spotify, Teamspeak, Twitch, Twitter, Waze, Whatsapp, Youtube, etc. As mentioned in section 2, the goal of this research is to classify traffic into application categories rather than identify individual applications so that the QoS requirements of the application categories can be identified and therefore accordingly configured by the SD-WAN network. Due to this reason, the application sessions from the trace were categorised into eight application categories and labelled accordingly. These categories are audio, browsing, chat, file transfer, mail, peer-to-peer (p2p), video, and VoIP. For each of these categories, a summary of their key characteristics and example of applications used in the trace are presented in table 2.

TABLE 2 TRAFFIC CATEGORIES WITH EXAMPLES OF APPLICATIONS USED IN THE TRAFFIC TRACE

Traffic Categories	*QoS Characteristics and trace applications used in the dataset*
Audio	Continuous stream of variable bit rate traffic, sensitive to delay and jitter. Session duration is generally long and communication is mostly non-interactive unidirectional. Traffic profile is consistent and follows a periodic pattern throughout the session. Example: Streaming audio services and podcast services from Spotify, iTunes, etc.
Browsing	Non-realtime, sensitive to loss, but can tolerate delay and jitter. Data transfer is in short burst and most of the transmission is downstream from the server to the client side. Session duration is limited to the roundtrip time of the request-reply process. Example: Traces include HTTP communication while browsing sites such as Facebook, Instagram, Google Maps, etc.
Chat	Semi-realtime, sensitive to loss, but can tolerate some delay. Transmission is in short burst and communication is bidirectional. Examples: Whatsapp, Facebook Messenger (MSN), etc.
File Transfer	Non-realtime traffic, sensitive to loss, but can tolerate transmission delay. Bursty unidirectional transmission with generally large burst sizes depending on available bandwidth. Transfer time is dependent on the size of the media. Example: File transfer to and from cloud platform such as iCloud, Dropbox, One Drive, etc
Mail	Non-realtime traffic, sensitive to loss, but can tolerate delay. Transmission is generally in small burst based on available bandwidth Example: Gmail, Yahoo, etc.
Peer-to-peer(P2P)	Traditionally non-realtime such as social file sharing services. Communication is generally bidirectional as peers may both upload and download simultaneously with other peers. The session durations are typically longer than mail or file transfer sessions.
Video	Variable bit rate and is sensitive to delay and jitter. Transmission period is generally longer than audio and requires more bandwidth. Traffic profile is consistent throughout the transmission session and follows a periodic pattern. Communication is mostly unidirectional in streaming services. Example: Youtube, Netflix, Twitch, etc.
VoIP	Variable bit rate interactive bi-directional traffic, sensitive to delay and jitter. Requires lower bandwidth than video Example: Teamspeak, Skype, etc.

4.2 Feature Selection

The goal of the feature selection stage is to find the smallest necessary set of statistically relevant features from the dataset (Kira and Rendell, 1992) that are required to achieve accuracy and precision in the prediction process. Irrelevant and redundant features increase the processing overhead without improving the accuracy. There has to be trade-off in the selection process to manage between high bias and high variance in the outcome. High bias will lead to loose generalisation in the mapping of the relationship resulting in high number of sessions being incorrectly grouped under the wrong category (under-fitting), whereas high variance will try to over-fit the dataset to the model losing the flexibility and generalisation resulting in missing sessions that do not fit to the strict profile. There are three main categories of feature selection. These are the wrapping methods, filter methods and embedded methods. In wrapping methods, different combination of subsets of the features are tested to find the most optimal subset. These methods are therefore computationally intensive and ML model dependent. Filter methods shortlist feature based on their descriptive semantics, variance of the feature and co-relation in-between the features. Hence, this category of schemes are independent of any specific ML model. In embedded methods, feature selection works in parallel with model creation. This involve adjustment of feature weights as the model evolves.

The original kaggle dataset had 87 features. This was first carefully filtered down to 29 temporal and spatial features based on their characteristics. Past research (Basu et al., 2002) have shown these shortlisted features to be correlated with the application profile and their QoS behaviour and are therefore relevant in estimating the traffic category. These first level shortlisted features are shown in table 3. For each session, these includes the five field tuple session identifiers comprising of the source and destination transport port numbers, protocol identifier and network layer addresses; and a range of temporal characteristics such as the duration of the session, a distribution comprising of the mean, standard deviation and the maximum and minimum value for the interarrival time of the overall bidirectional session and the individual unidirectional flows within it and the similar distribution for the active and idle period of the flow.

TABLE 3 FIRST-LEVEL SHORTLISTED FEATURES

Source and destination Port
Transport layer protocol
Source and destination IP
Session duration
Session size in bytes and packet count
Mean, standard deviation, maximum and minimum interarrival time of the overall session
Mean, standard deviation, maximum and minimum interarrival time for the forward and reverse flows
Mean, maximum and standard deviation of the active and idle period of the session

At the next stage, these shortlisted features are further narrowed down by analysing the performance of different subset of these features on its predictive ability as well as the correlation and redundancy between them. The Waikato Environment for Knowledge Analysis (WEKA) (Srivastava, 2014) ML library's Correlation based Feature Selection Subset Evaluation algorithm (CfsSubsetEval) along with the best first search method was used for this purpose.

For cross validation of the impact of the feature selection algorithm on the performance of the ML models, a relevant second feature selection model from the WEKA library called Gain Ratio Attribute Evaluation was used. This model selects attributes by shortlisting the subset of attributes that results in the most information gain (Priyadarsini et al., 2011). The shortlisted features from these two algorithms are listed in table 4.

TABLE 4 SHORTLISTED FEATURES FROM THE FEATURE SELECTION ALGORITHMS

Feature Selection Algorithms	
Correlation based	***Gain Ratio based***
Source Port	Source IP
Destination Port	Destination IP
Session duration	Source Port
Session rate	Destination Port
Reverse flow's interarrival time	Session rate

5 PERFORMANCE METRICS FOR COMPARISON

The goal of the performance metrics in this research is to facilitate transparent validation and comparison of the performance of the ML models in terms of its accuracy of prediction, robustness of the models and flexibility in dealing with a wide range of Internet traffic categories. Since the focus here is on classification based ML models (rather than regression models), the relevant metrics need to provide some sort of proportional measurement of the accurately of the models in terms of the number of correct or incorrect predictions in proportion to the total number of instances in that category (Seliya et al., 2009). In this context, the following confusion matrix (Fig. 4) and the associated measurement metrics derived from it are suitable for this research:

		Actual Outcome	
		Positive	***Negative***
Predicted Outcome	***Positive***	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

Figure 4 Confusion matrix

True Positive Rate (TPR)

This measurement is the ratio of the total number of traffic sessions correctly classified as belonging to a category (True Positive TP) in proportion to the total number of actual sessions belonging to the category, which includes the TP sessions and the sessions incorrectly classified as not belonging to the category (False Negative FN). This is expressed as:

$$TPR = \frac{TP}{TP + FN} \qquad (1)$$

In literature, this metric is also referred as recall, detection rate (DR) or sensitivity. Note that in this multi-class classification scenario for classifying different traffic category, for each class of traffic a separate one-vs-rest binary classification is first required and then the results are convoluted.

False Positive Rate (FPR)

The FPR is the ratio of the number of sessions incorrectly classified as belonging to the category (False Positive FP) to the total number of actual sessions that do not belong to the category, which includes FP along with the number of sessions correctly classified as not belonging to the category (True Negative TN). This is the opposite of TPR and is expressed as:

$$FPR = \frac{FP}{FP + TN} \qquad (2)$$

Received Operating Characteristics (ROC) curve

The ROC curve represents a graphical plot of TPR along the y-axis vs FPR along the x-axis for different values of the classification threshold and is useful for optimising the performance of a ML model. A good model will result in a ROC curve with high gradient and the area under the ROC curve close to 1.

Precision

Precision is the ratio of the number of sessions correctly predicted as belonging to a traffic category (TP) to the total number of sessions that were predicted as belonging to that category including the sessions incorrectly predicted as belonging to the category. This is represented as:

$$Precision = \frac{TP}{TP + FP} \qquad (3)$$

F-Measure

The TPR value gives the proportion of the traffic sessions correctly predicted whereas precision shows the confidence level in the positive prediction process. F-Measure provides the comparative compromise between these two measurements and is represented as:

$$F - Measure = \frac{Precision\ x\ TPR}{Precision + TPR} \qquad (4)$$

The value of F-Measure could be used to analyse the effectiveness of a particular ML model with a specific feature selection algorithm and the ideal value of this metric is 1.

Accuracy

The accuracy value provides the most generalised measurement and is the ratio of the number of correct predictions (both TP and TN) to the total number of predictions. It can be represented as:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \qquad (5)$$

It gives a general overview of the performance of a ML model which can then be further analysed with the more specific metrics (discussed above) derived from the confusion matrix.

In the next section, we present a study of the performance of the MLs models using these metrics as the comparison parameters.

6 ANALYSIS OF THE PERFORMANCE OF THE MACHINE LEARNING MODELS

In our scenario, for traffic category prediction, the traffic categories are known in advance as well as training and testing data can be readily made available to test and tune the performance of the models with a wide range of suitable features. In this situation, supervised machine learning models will provide the best performance because of the inherent nature of these types of models to work on predefined classes and structured set of training and testing data. If the traffic classes were unknown, we could have considered the unsupervised models for clustering the traces for further analysis (Usama et al., 2019). For unknown or multidimensional complex feature extraction, deep learning would have also been an alternative (Wang et al., 2020). However, here the features are one dimensional and well-defined. Hence, there will be no additional benefits for the comparative extra resource and time of using deep learning models. Hence, in our proposed SD-WAN framework, we have focussed on supervised ML models for the prediction module.

The shortlisted features for the ML models in both table 3 and table 4 consist of both range based continuous numeric data such as session rate and interarrival time as well as discrete pattern matching data such as port id and IP addresses. These features can be either readily extracted or easily computed from the arriving packets. We trained, analysed and compared the performance of a range of ML models with both the first level of shortlisted features (from table 3) as well as with the features further selected by the feature selection algorithms (table 4).

Naïve Bayes, Support Vector Machine (SVM) and Logical Regression

These three ML models (Moore and Zuev, 2005; Dong, 2021) showed the weakest performance in predicting the traffic categories of the sessions with all the three sets of features. The accuracy rate of the three models is shown in table 5.

TABLE 5 ACCURACY OF THE NAÏVE BAYES, SVM AND LOGICAL REGRESSION MODELS

	Feature Selection Algorithms		
	1st level shortlist	***Correlation based***	***Gain Ratio based***
Naïve Bayes	38.43%	43.51%	52.94%
SVM	80.19%	55.92%	72.67%
Logical Regression	87.86%	66.6%	80.9%

Similar results were also seen with the other comparison metrics across different traffic categories. Naïve Bayes is dependent on the profile and distribution of the training dataset and its performance can vary significantly if the test set differ from the distribution as is evident from the results. This would also make Naïve Bayes unsuitable for realtime traffic classification in the SDN context. The results from the SVM model are comparatively better, but are still unacceptable. For classification using both linear and higher order polynomial kernel, the improvement was minimal. The model performed best when all the 29 first-level features were used for classification. Logistic Regression gave the best result among the three models across all the three subsets of features. The inherent binary classification model was extended to multinomial logistic regression to support multiclass classification. To balance between over and under-fitting, the ridge value was set to 1.0e-8 in the experiments. Hence, these models can be eliminated as contenders for our traffic classification scenario.

Instance Based Learner (IBK)

This scheme is more robust than the previous models and it uses the k most exact training patterns to make the prediction (MeeraGandhi, 2010). In the experiments, the value of k was set at runtime using cross validation. The distance measurement for grouping neighbours was done based on Euclidean distance since the features were mainly numeric data. In the results, the highest accuracy rate was 99.5% using the subset of features derived from the gain ratio based feature selection scheme. The results from the validation and the corresponding confusion matrix for each class of traffic category are shown in tables 6 and 7.

TABLE 6 PERFORMANCE OF IBK ON EACH CLASS OF TRAFFIC USING GAIN RATIO BASED ATTRIBUTE EVALUATION

Class	TP Rate	FP Rate	Precision	F-Measure	ROC Area
Audio	0.997	0.002	0.984	0.99	0.999
Browsing	0.993	0.002	0.996	0.994	0.996
Chat	0.978	0	0.994	0.986	0.988
File Trans	0.992	0	0.998	0.995	0.996
Mail	1	0	0.993	0.996	1
P2P	0.994	0	0.997	0.996	0.997
Video	0.995	0.001	0.992	0.994	0.996
VoIP	0.999	0.001	0.998	0.998	0.999
Weighted Avg.	**0.995**	**0.001**	**0.995**	**0.995**	**0.997**

TABLE 7 CONFUSION MATRIX OF IBK FOR EACH CLASS OF TRAFFIC USING GAIN RATIO BASED ATTRIBUTE EVALUATION

Classed As -->	a	b	c	d	e	f	g	h
a = Audio	719	1	0	0	0	1	0	0
b = Browsing	6	1592	2	0	0	2	2	0
c = Chat	2	1	316	0	2	0	1	1
d =File Trans	2	0	0	857	0	0	3	2
e = Mail	0	0	0	0	282	0	0	0
f = P2P	2	4	0	0	0	1079	0	0
g = Video	0	1	0	1	0	0	870	2
h = VoIP	0	0	0	1	0	0	1	2289

Results from only first-level feature selection had a slightly lower accuracy of 93.27% and are presented in tables 8 and 9.

TABLE 8 PERFORMANCE OF IBK ON EACH CLASSOF TRAFFIC USING ONLY 1ST-LEVEL SHORTLISTED ATTRIBUTES

Class	TP Rate	FP Rate	Precision	F-Measure	ROC Area
Audio	0.85	0.017	0.831	0.84	0.918
Browsing	0.89	0.025	0.898	0.894	0.936
Chat	0.802	0.006	0.852	0.826	0.907
File Trans	0.957	0.003	0.974	0.966	0.978
Mail	0.883	0.005	0.862	0.872	0.948
P2P	0.982	0.002	0.985	0.983	0.989
Video	0.923	0.014	0.892	0.907	0.957
VoIP	0.984	0.006	0.986	0.985	0.989
Weighted Avg.	**0.993**	**0.011**	**0.993**	**0.933**	**0.933**

TABLE 9 CONFUSION MATRIX OF IBK FOR EACH CLASS OF TRAFFIC USING ONLY 1ST-LEVEL SHORTLISTED ATTRIBUTES.

Classed As -->	a	b	c	d	e	f	g	h
a = Audio	613	80	8	11	0	9	0	0
b = Browsing	99	1428	11	0	3	6	48	9
c = Chat	9	25	259	0	3	1	19	7
d =File Trans	10	4	1	827	11	0	7	4
e = Mail	0	4	2	7	249	0	13	7
f = P2P	7	11	1	0	1	1065	0	0
g = Video	0	32	12	1	16	0	807	6
h = VoIP	0	6	10	3	6	0	11	2255

Nearest neighbor with generalization (NNge)

It is another type of instance-based learning model that uses case-based reasoning using generalised exemplars (Carela-Español, 2010). It creates if-then-else rule for defining the grouping for the discrete and continuous features and their distances. The NNge model gave the best accuracy with the gain ratio based attributes (99.68%). The results from the validation run are shown in tables 10 and 11.

TABLE 10 PERFORMANCE OF NNGE ON EACH CLASSOF TRAFFIC USING GAIN RATIO BASED ATTRIBUTE EVALUATION

Class	TP Rate	FP Rate	Precision	F-Measure	ROC Area
Audio	0.996	0.001	0.994	0.995	0.998
Browsing	0.997	0.001	0.998	0.997	0.998
Chat	0.985	0.001	0.981	0.985	0.992
File Trans	0.999	0	0.999	0.999	0.999
Mail	1	0	0.993	1	1
P2P	0.996	0.001	0.996	0.996	0.998
Video	0.998	0	1	0.998	0.999
VoIP	0.998	0.001	0.998	0.998	0.999
Weighted Avg.	**0.997**	**0.001**	**0.997**	**0.997**	**0.998**

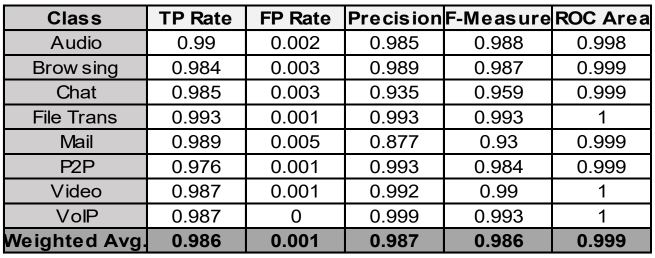
TABLE 11 CONFUSION MATRIX OF NNGE FOR EACH CLASS OF TRAFFIC USING GAIN RATIO BASED ATTRIBUTE EVALUATION

Classed As -->	a	b	c	d	e	f	g	h
a = Audio	718	2	0	0	0	1	0	0
b = Browsing	1	1599	1	0	0	2	0	1
c = Chat	0	0	318	1	2	1	0	1
d =File Trans	0	0	0	863	0	0	0	1
e = Mail	0	0	0	0	282	0	0	0
f = P2P	3	1	0	0	0	1081	0	0
g = Video	0	0	1	0	0	0	872	1
h = VoIP	0	1	4	0	0	0	0	2286

Bayseian Network (Bayes Net)

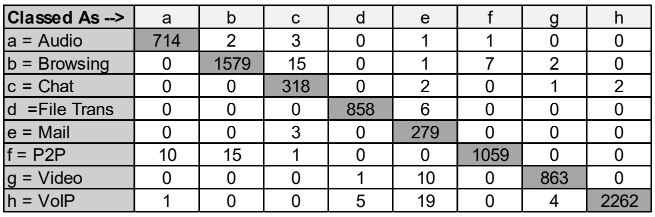
It presents a directed acyclic graph representation of the conditional probabilities of the features which are assumed to be discrete and finite (Moore and Zuev, 2005). Since the conditional dependencies of the features are limited in our case, the accuracy rate from the Bayes Net model is comparatively lower than the NNge. Its best result was 98.6% with the gain ration based attributes. The validation results and the corresponding confusion matrix are shown in tables 12 and 13 respectively.

TABLE 12 PERFORMANCE OF BAYES NET ON EACH CLASSOF TRAFFIC USING GAIN RATIO BASED ATTRIBUTE EVALUATION



Class	TP Rate	FP Rate	Precision	F-Measure	ROC Area
Audio	0.99	0.002	0.985	0.988	0.998
Brow sing	0.984	0.003	0.989	0.987	0.999
Chat	0.985	0.003	0.935	0.959	0.999
File Trans	0.993	0.001	0.993	0.993	1
Mail	0.989	0.005	0.877	0.93	0.999
P2P	0.976	0.001	0.993	0.984	0.999
Video	0.987	0.001	0.992	0.99	1
VoIP	0.987	0	0.999	0.993	1
Weighted Avg.	**0.986**	**0.001**	**0.987**	**0.986**	**0.999**

TABLE 13 CONFUSION MATRIX OF BAYES NET FOR EACH CLASS OF TRAFFIC USING GAIN RATIO BASED ATTRIBUTE EVALUATION

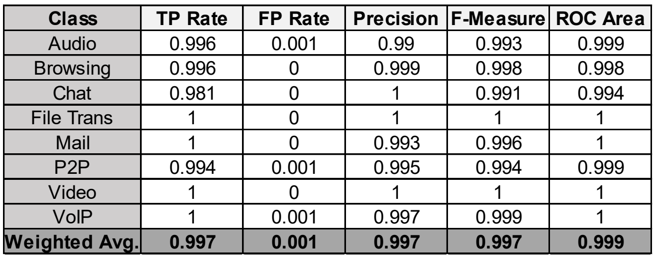


Classed As -->	a	b	c	d	e	f	g	h
a = Audio	714	2	3	0	1	1	0	0
b = Browsing	0	1579	15	0	1	7	2	0
c = Chat	0	0	318	0	2	0	1	2
d =File Trans	0	0	0	858	6	0	0	0
e = Mail	0	0	3	0	279	0	0	0
f = P2P	10	15	1	0	0	1059	0	0
g = Video	0	0	0	1	10	0	863	0
h = VoIP	1	0	0	5	19	0	4	2262

J48 Decision tree

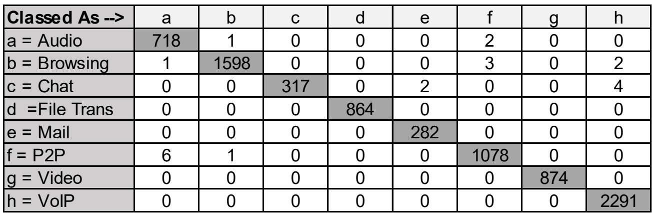
It builds a top-down, recursive decision tree by selecting the most appropriate attributes from the features at each split point (Bakhshi and Ghita, 2016). The attribute at a split point is selected by calculating the information gain for each of the available attributes at the split point and then selecting the attribute that gives the maximum gain. The J48 decision tree can handle both continuous and discrete values and is ideal for network traffic features, which is composed of both of these types of attributes. The results from this model with gain ratio based attribute selection are presented in tables 13 and 14 respectively. This shows the best accuracy rate (99.72%) of all the combination of ML models and feature selection algorithms and provides the ideal trade-off between over and under fitting of the data. The TP rate, precision and the area under the ROC curve either is very close to the ideal value of 1 or is 1 itself for most traffic categories. This is also evident in the confusion matrix where the number of incorrect predictions is low.

TABLE 14 PERFORMANCE OF J48 ON EACH CLASSOF TRAFFIC USING GAIN RATIO BASED ATTRIBUTE EVALUATION



Class	TP Rate	FP Rate	Precision	F-Measure	ROC Area
Audio	0.996	0.001	0.99	0.993	0.999
Browsing	0.996	0	0.999	0.998	0.998
Chat	0.981	0	1	0.991	0.994
File Trans	1	0	1	1	1
Mail	1	0	0.993	0.996	1
P2P	0.994	0.001	0.995	0.994	0.999
Video	1	0	1	1	1
VoIP	1	0.001	0.997	0.999	1
Weighted Avg.	**0.997**	**0.001**	**0.997**	**0.997**	**0.999**

TABLE 15 CONFUSION MATRIX OF J48 DECISION TREE FOR EACH CLASS OF TRAFFIC USING GAIN RATIO BASED ATTRIBUTE EVALUATION



Classed As -->	a	b	c	d	e	f	g	h
a = Audio	718	1	0	0	0	2	0	0
b = Browsing	1	1598	0	0	0	3	0	2
c = Chat	0	0	317	0	2	0	0	4
d =File Trans	0	0	0	864	0	0	0	0
e = Mail	0	0	0	0	282	0	0	0
f = P2P	6	1	0	0	0	1078	0	0
g = Video	0	0	0	0	0	0	874	0
h = VoIP	0	0	0	0	0	0	0	2291

7 EVALATION

The dataset of the traffic traces used in this research covered a wide spectrum of applications from realtime video and VoIP to non-realtime web browsing and file transfer and captures the holistic heterogeneity of modern Internet traffic categories. Based on the analysis in the previous section, the precision of different application categories with different ML models is summarised in figure 5. As discussed in equation 3, the precision provides the accuracy of the positive prediction rate and hence is an important benchmark for selecting the appropriate ML model for identification of the traffic category of the incoming flow. It can be seen that the P2P applications provide the most consistent precision value across the different ML models. Again, the J48 model gives the highest precision values across all categories of applications.

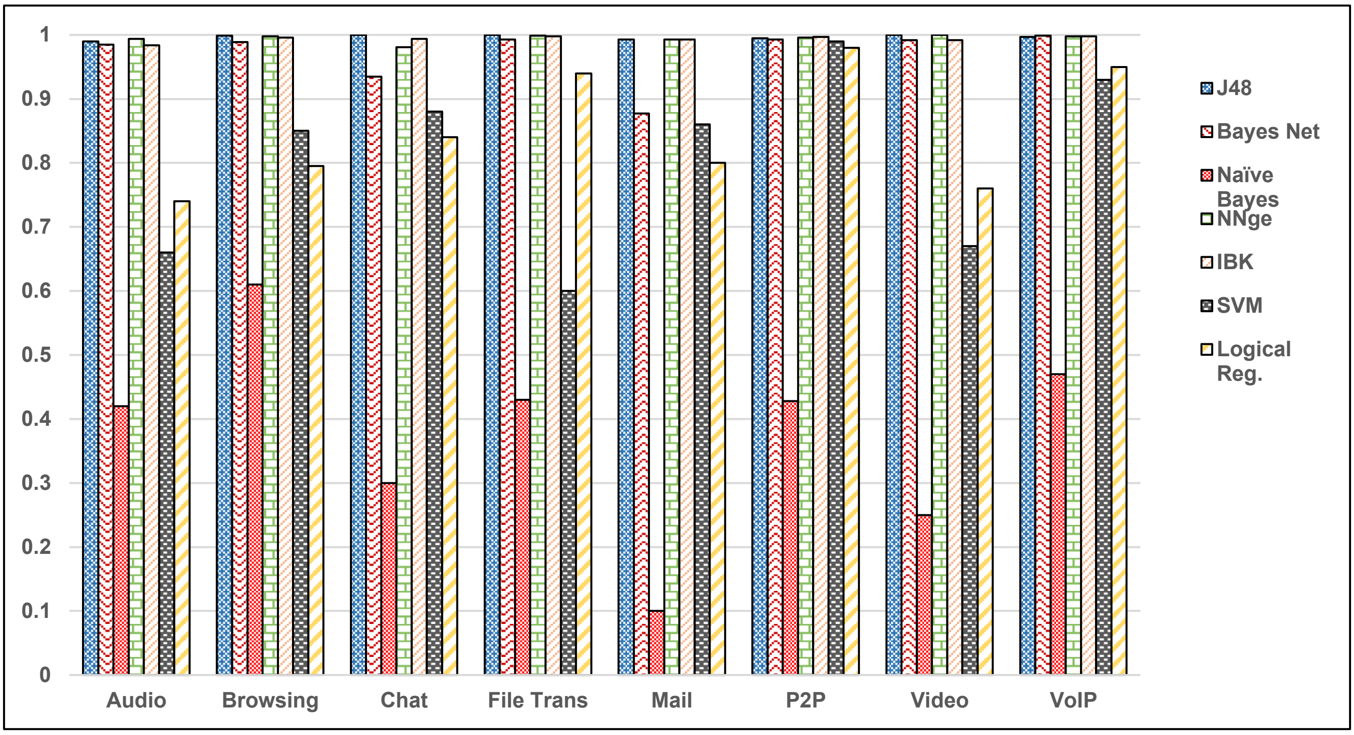


Figure. 5 Precision of the ML models for different categoris of applications

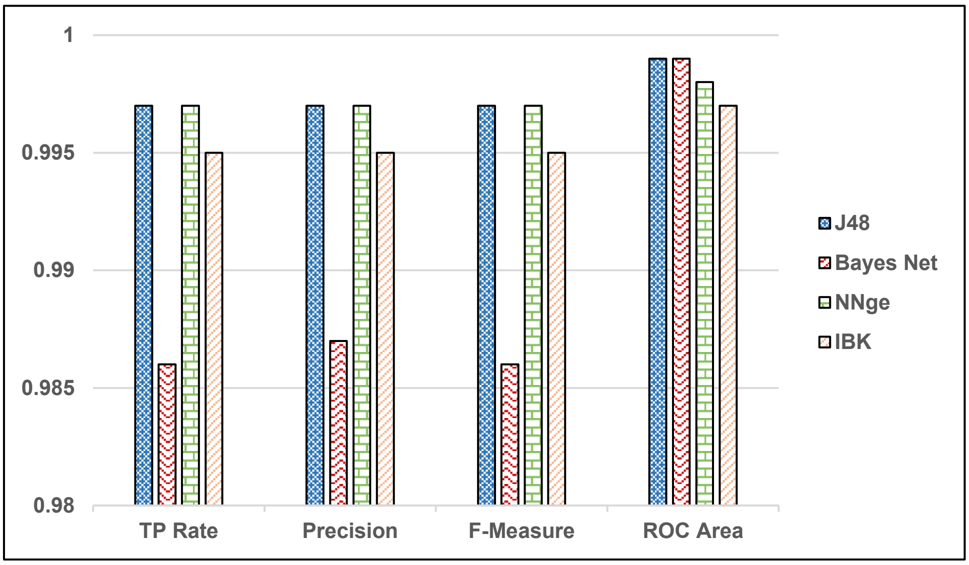


Figure. 6 Performance of the ML models with the aggregated traffic

A holistic analysis of the performance of the ML models across the weighted average of the aggregated traffic categories when compared against the key performance metrics is presented in figure 6. In this case too, the J48 shows the best performance among the shortlisted models.

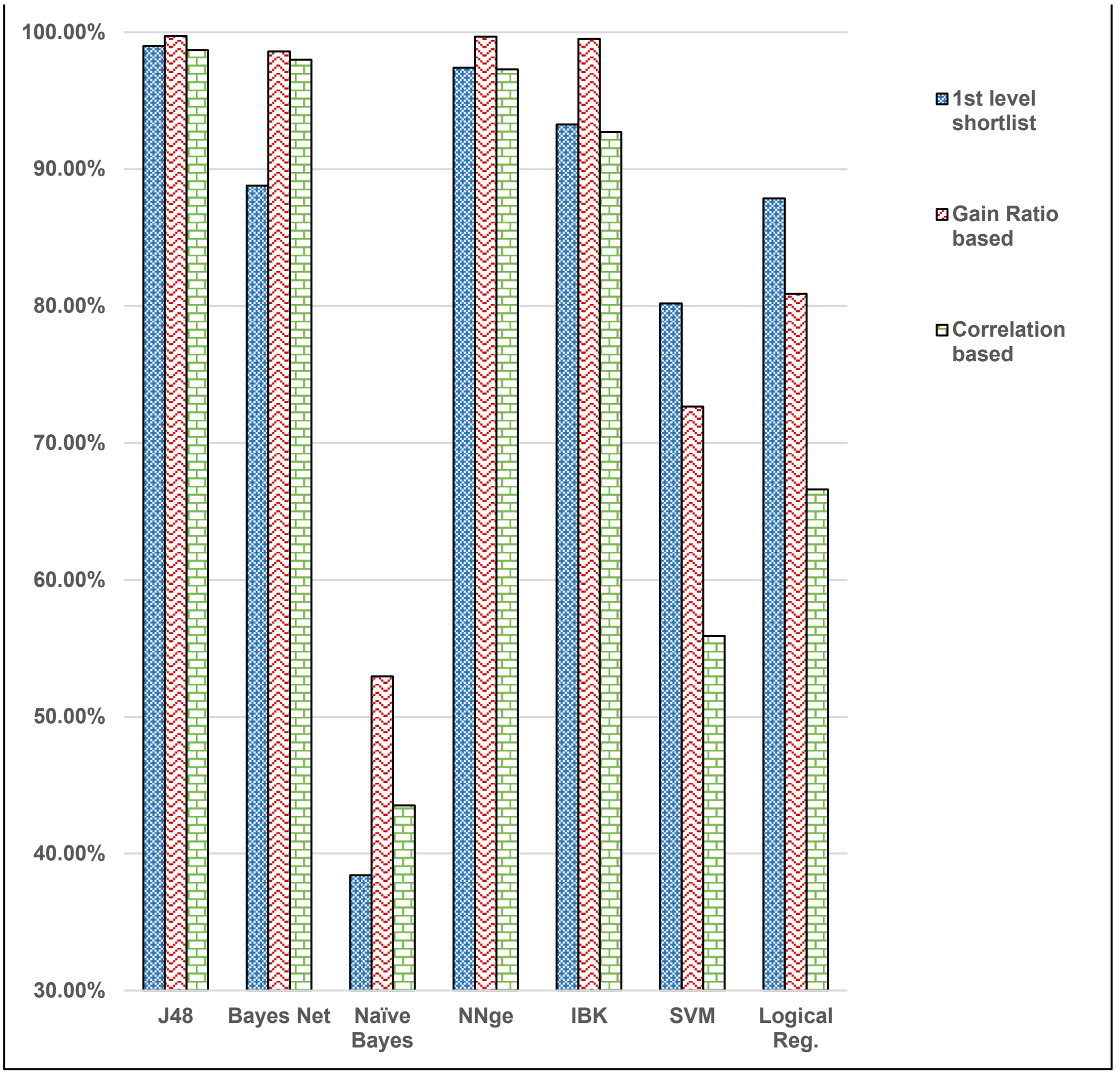


Figure 7. Accuracy of the models with different attribute selection algorithms

The accuracy rate of a model shows the rate of correct prediction (Equation 5). The impact of the feature selection schemes on the accuracy is shown in figure 7. As discussed in section 5, it is evident from the figure that the gain ratio based attribute selection scheme gives the best results for all the models except for the SVM and logical regression models. These models also had a comparatively higher error rate compared to the other models for all the performance metrics.

It can be concluded from the research that the combination of J48 decision tree ML model with the gain ratio based attribute selection scheme is the most suitable prediction scheme for Internet traffic and gives a 99.72% accuracy. The results from the NNge model with gain ratio based attributes is quite close with 99.68% accuracy. Hence, both of these models could be used for prediction purpose and could act as a benchmark to compare the performances of other models.

8 CONCLUSION

This research extends the SD-WAN model by proposing a novel architecture of a scalable SD-WAN framework that can incorporate different types of AI/ML centric network applications in its application plane. In this research, we have designed and developed a ML-based traffic prediction module for the traffic engineering unit of the application plane and have shown its relationship and role in the context of the overall SD-WAN. The correctness of the prediction process is significant for the SD-WAN as all follow-on actions of the traffic engineering unit and other applications in the application plane is based on this outcome. For the traffic prediction module, we have analysed the performance of a set of seven relevant supervised ML models with actual traffic traces from a wide range of Internet traffic categories and with different subset of features selected methodically using feature selection techniques. Based on the analysis, we have found that the J48 ML model and the associated gain ratio based attributes gives the best prediction rate for the arriving traffic at the ingress point of the SD-WAN network.

REFERENCES

Akyildiz, I.F., Anjali, T., Chen, L., de Oliveira, J.C., Scoglio, C., Sciuto, A., Smith J.A. and Uhl, G. (2003), "A new traffic engineering manager for DiffServ/MPLS networks: design and implementation on an IP QoS Testbed," *Computer Communications*, Vol. 26 No. 4, pp. 388-403.

Azab, A., Khasawneh, M., Alrabaee, S., Choo, K.K.R. and Sarsour, M. (2022), "Network traffic classification: Techniques, datasets, and challenges," *Digital Communications and Networks*.

Bakhshi, T. and Ghita, B. (2016), "On Internet Traffic Classification: A Two-Phased Machine Learning Approach," *Journal of Computer Networks and Communications*, Hindawi, Vol. 2016: 2048302.

Basu, K., Ball F. and Kouvatsos, D.D. (2002), "A Simulation Study of IPV6 to ATM Flow Mapping Techniques". *SCS Transaction Journal on Network Modeling and Performance Issues*, Vol. 78 No. 7, pp. 423-430.

Basu, K., Hamdullah, A. and Ball, F. (2020), "Architecture of a Cloud-based Fault-Tolerant Control Platform for improving the QoS of Social Multimedia Applications on SD-WAN*," The 13th IEEE International Conference on Communications (COMM2020)*, Bucharest, Hungary, June 18-20, 2020.

Bogle, J., Bhatia, N., Ghobadi, M., Menache, I., Bjørner, N., Valadarsky, A. and Schapira, M. (2019), "TEAVAR: striking the right utilization-availability balance in WAN traffic engineering," *Proceedings of the ACM Special Interest Group on Data Communication*, pp. 29-43.

Boutaba, R., Salahuddin, MA., Limam, N., Ayoubi, S., Shahriar, N., Estrada-Solano, F. and Caicedo, O.M. (2018), "A comprehensive survey on machine learning for networking: evolution, applications and research opportunities," *Journal of Internet Services and Applications*, Vol. 9 No. 1, pp. 16:1-16:99.

Bremler-Barr, A., Harchol, Y., Hay D. and Koral, Y. (2014), "Deep packet inspection as a service," *10th ACM International on Conference on emerging Networking Experiments and Technologies*, pp. 271-282.

Bujlow, T., Carela-Español, V. and Barlet-Ros, P. (2015), "Independent comparison of popular DPI tools for traffic classification," *Computer Networks*, Vol. 76, pp.75-89.

Carela-Español, V., Barlet-Ros, P., Solé-Simó, M., Dainotti, A., de Donato, W. and Pescapé, A. (2010), "K-Dimensional Trees for Continuous Traffic Classification," in Traffic Monitoring and Analysis. *TMA 2010. Lecture Notes in Computer Science*, Ricciato, F., Mellia, M., Biersack, E. (eds), Springer, Vol. 6003.

Dong, S. (2021), "Multi class SVM algorithm with active learning for network traffic classification," *Expert Systems with Applications*, Vol. 176 No. 114885.

Ghosh, A. and Senthilrajan, A., 2019. Classifying network traffic using dpi and dfi. International journal of scientific and technology research, 8(11), p.1019.

Ghosh, D., Sarangan, V. and Acharya, R. (2001), "Quality-of-service routing in IP networks," in *IEEE Transactions on Multimedia*, Vol. 3 No. 2, pp. 200-208.

Gringoli, F., Salgarelli, L., Dusi, M. , Cascarano, N. , Risso, F. and Claffy, K. C. (2009), "GT: picking up the truth from the ground for Internet traffic," *SIGCOMM Comput. Commun. Rev*, Vol. 39 No. 5, pp. 12-18.

Hu, F., Hao, Q. and Bao, K. (2014), "A survey on software-defined network and openflow: From concept to implementation," *IEEE Communications Surveys & Tutorials*, Vol. 16 No. 4, pp. 2181-2206.

Internet Assigned Numbers Authority:IANA (2023), *Service Name and Transport Protocol Port Number Registry*, https://www.iana.org/assignments/service-names-port-numbers/service-names-port-numbers.xhtml. Accessed 15 March 2023.

Jose, A. V., Selvan, M. P., Mary, V. A., Jancy, S., Helen, S. P, A. (2022), "Prediction of Network Attacks Using Supervised Machine Learning Algorithm," *2022 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, Chennai, India, pp. 1-5.

Khater, N. A. and R. E. Overill (2015), "Network traffic classification techniques and challenges," *10th International Conference on Digital Information Management (ICDIM)*, pp. 43-48,

Kira, K. and Rendell, L. A. (1992), "A Practical Approach to Feature Selection," *Machine Learning Proc.*, pp. 249-256.

Latif, Z., Sharif, K., Li, F., Karim, M.M., Biswas, S. and Wang, Y. (2020), "A comprehensive survey of interface protocols for software defined networks," *Journal of Network and Computer Applications*, Vol. 156, p.102563.

Liu, W. X., Cai, J. , Wang, Y. , Chen, Q. C. and Zeng, J. (2020), “Fine-grained flow classification using deep learning for software defined data center networks,” *Journal of Network and Computer Applications*, vol. 168:102766

Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T. and Wright, C. (2014), “Virtual extensible local area network (VXLAN): A framework for overlaying virtualized layer 2 networks over layer 3 networks,” *rfc7348*.

MeeraGandhi, G. (2010), “Machine Learning Approach for Attack Prediction and Classification using Supervised Learning Algorithms,” *International Journal of Computer Science & Communication*, Vol. 1 No. 2, pp. 247-250.

Moore, A.W., Papagiannaki, K. (2005), “Toward the Accurate Identification of Network Applications,” *International workshop on passive and active network measurement,* Berlin, Germany, pp. 41-54.

Moore, A. W. and Zuev, D. (2005), “Internet traffic classification using Bayesian analysis techniques,” ACM Sigmetrics *International Conference on Measurement and Modeling of Computer Systems*, pp. 50-60.

Priyadarsini, R. P., Valarmathi, M.L. and Sivakumari, S. (2011), “Gain Ration Feature Selection Method for Privacy Preservation,” *ICTACT Journal on Soft Computing*, Vol. 1 No. 4, pp. 201 – 205.

Rojas, J. S., Pekar, A. , Rendón Á. and Corrales J. C. (2020), "Smart User Consumption Profiling: Incremental Learning-Based OTT Service Degradation," *IEEE Access*, Vol. 8, pp. 207426-20744.

Rojas, J. S. (2017), *IP Network Traffic Flows Labeled with 75 Apps*, kaggle, available at: https://www.kaggle.com/datasets/jsrojas/ip-network-traffic-flows-labeled-with-87-apps?select=Dataset-Unicauca-Version2-87Atts.csv (accessed 19th Dec 2022).

Salman, O., Elhajj, I.H., Kayssi, A. and Chebab A. (2020), “A review on machine learning–based approaches for Internet traffic classification,” *Annals of Telecommunications,* Vol. 75, pp. 673–710.

Santiago Lopes Pereira, S. , De Castro e Silva, J. L. and Bessa Maia, J. E. (2014), "NTCS: A real time flow-based network traffic classification system," *10th International Conference on Network and Service Management (CNSM),* pp. 368-371.

Schueller, Q., Basu K., Younas, M., Patel M. and Ball F. (2018), “A Hierarchical Intrusion Detection System using Support Vector Machine for SDN Network in Cloud Data Center,” *28th IEEE International Telecommunication Networks and Applications Conference (ITNAC 2018)*, Sydney, Australia.

Seliya, N., Khoshgoftaar T. M. and Van Hulse, J. (2009) , “A Study on the Relationships of Classifier Performance Metrics,” *21st IEEE International Conference on Tools with Artificial Intelligence*, pp. 59-66.

Srivastava, S. (2014), “Weka: A tool for data preprocessing, classification, ensemble, clustering and association rule mining,” *International Journal of Computer Applications*, Vol. 88 No. 10.

Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K.L.A., Elkhatib, Y., Hussain, A. and Al-Fuqaha, A. (2019), “Unsupervised machine learning for networking: Techniques, applications and research challenges,” *IEEE Access*, Vol. 7, pp. 65579-65615.

Xue, Y., Wang, D. and Zhang, L. (2013), “Traffic classification: Issues and challenges,” *International Conference on Computing, Networking and Communications (ICNC)*, San Diego, CA, USA, pp. 545-549

Wang, X., Ma, Y., Wang, Y., Jin, W., Wang, X., Tang, J., Jia, C. and Yu, J. (2020), “Traffic flow prediction via spatial temporal graph neural network,” *In Proceedings of the web conference* 2020, pp. 1082-1092.

Wang, Z. (2015), “The applications of deep learning on traffic identification,” BlackHat USA, Vol. 24 No. 11, pp.1-10.

Yang, Z., Cui, Y., Li, B., Liu, Y. and Y. Xu (2019), “*Software-Defined Wide Area Network (SD-WAN): Architecture, Advances and Opportunities,*” 28th International Conference on Computer Communication and Networks (ICCCN), Valencia, Spain, pp. 1-9.