

# Analysis of Artificial Intelligence Methods for Automatic Bandwidth Adjustment for Wireless Networks

Carrillo Marlon<sup>1</sup>, Torres Rommel<sup>2</sup>, Barba Luis<sup>3</sup>

**Abstract** — The exponential increase in Internet traffic is mainly due to the proliferation of services such as audio and video streaming, the emergence of applications that require a lot of bandwidth to work optimally and generally the process of digitalization of services. In this context, bandwidth management plays a fundamental role, which translates into a better experience for users. Traffic congestion causes the exchange of information to become deficient, that is why techniques such as automatic bandwidth adjustment have been investigated, which manages the bandwidth according to the traffic demand, therefore in this document a study is made about the automatic bandwidth adjustment, the way in which Artificial Intelligence is integrated with computer networks, finally a comparison will be made of several machine learning methods, cataloged within supervised learning, carrying out several experiments determining that Random Forest is the most effective method to predict the automatic bandwidth adjustment, followed by Naive Bayes, Logistic Regression, and Support Vectorial Machine (SMV), on the other hand K -nearest neighbor (KNN) and neural network do not demonstrate considerable effectiveness, each experiment was carried out taking into account the Quality of Service (QoS).

**Keywords:** traffic, Internet, Machine Learning, Supervised Learning, Quality of Service.

**Resumen** — El aumento exponencial del tráfico en Internet se debe principalmente a la proliferación de servicios como el streaming de audio y video, a la aparición de aplicaciones que requieren un alto ancho de banda para funcionar de manera óptima, y, en general, al proceso de digitalización de servicios. En este contexto, la gestión del ancho de banda desempeña un papel fundamental, ya que contribuye a una mejor experiencia para los usuarios. La congestión de tráfico genera que el intercambio de información sea ineficiente. Por ello, se han investigado técnicas como el ajuste automático de ancho de banda, que permite gestionar el ancho de banda según la demanda de tráfico. En este documento se realiza un estudio sobre el ajuste automático de ancho de banda, abordando cómo la Inteligencia Artificial se integra con las redes de computadoras. Además, se presenta una comparación de varios métodos de aprendizaje supervisado en machine learning. A través de diversos experimentos, se determinó que Random Forest es el método más efectivo para predecir el ajuste automático de ancho de banda, seguido por Naive Bayes, Regresión Logística y Support Vector Machine (SVM). En contraste, los métodos K-Vecinos Más Cercanos (KNN) y redes neuronales no demostraron una efectividad significativa. Cada experimento se realizó considerando la Calidad de Servicio (QoS).

**Palabras Clave:** tráfico, Internet, Machine Learning, Aprendizaje Supervisado, Calidad de Servicio.

---

This paragraph of the first footnote will contain the date on which you submitted your paper for review, which is populated by IEEE. It is IEEE style to display support information, including sponsor and financial support acknowledgment, here and not in an acknowledgment section at the end of the article. For example, "This work was supported in part by the U.S. Department of Commerce under Grant 123456." The name of the corresponding author appears after the financial information, e.g. (*Corresponding author: Second B. Author*). Here you may also indicate if authors contributed equally or if there are co-first authors.

The next few paragraphs should contain the authors' current affiliations, including current address and e-mail. For example,

1. Carrillo Encalada Marlon Vicente, is an Information Technology Engineer, graduated from the Universidad Técnica Particular de Loja (e-mail: [mv-carrillo2@utpl.edu.ec](mailto:mv-carrillo2@utpl.edu.ec)). ORCID: 0009-0005-1289-3867

2. Torres Tandazo Rommel Vicente, is Torres Rommel, holds a PhD in Computer Science from the Polytechnic University of Madrid, and is currently a research professor in the Department of Computer Science and Electronics at the Technical University of Loja. (e-mail: [rovitor@utpl.edu.ec](mailto:rovitor@utpl.edu.ec)). ORCID: 0000-0003-2313-0118

3. Barba Guamán Luis Rodrigo, holds a PhD in Science and Technology for Smart Citys from the Polytechnic University of Madrid, and is currently a research professor in the Department of Computer Science and Electronics at the Technical University of Loja, (e-mail: [lrbarba@utpl.edu.ec](mailto:lrbarba@utpl.edu.ec)). ORCID: 0000-0002-8569-3322

Manuscript Received: 16/10/2024

Revised: 07/12/2024

Accepted: 15/12/2024

DOI: <https://doi.org/10.29019/enfoqueute.1105>

Section Editor: Marcelo Mosquera

## I. INTRODUCTION

TECHNOLOGY has advanced significantly in recent years, today we depend on the Internet for everyday actions such as communication, entertainment, information search, digital services and others. Bandwidth can be defined as: the amount of information that is transmitted through a network connection in each period. It is crucial for proper Internet browsing, because the greater the amount of information transmitted in less time, the better the user experience, for the proper functioning of the services, the availability of bandwidth is crucial [1].

The exponential growth of services such as audio and video streaming, the proliferation of applications that consume a lot of bandwidth, the significant growth of online services and the increase of smart devices that integrate the Internet of Things (IoT), has caused the constant updating of control measures to ensure an adequate quality of service (QoS) to users. The amount of information transmitted on the Internet has gone from bps to Gbps, due to this it is necessary to make changes and implement measures to offer a better service to users [2].

In this research, the integration of Artificial Intelligence methods and algorithms for efficient distribution for bandwidth management is proposed. The objectives of our approach are:

Obtain a set of metrics for bandwidth requirements related to quality of service.

Define scenarios for the generation of dataset information or collect datasets for the use of artificial intelligence tools.

Analysis of artificial intelligence (AI) techniques related to automatic bandwidth adjustment.

Compare AI techniques using QoS requirements for dynamic bandwidth adjustment.

This research is an initial work, its findings will be used for the creation of a wireless network automatic bandwidth adjustment model based on Artificial Intelligence methods with the capacity to predict the required bandwidth.

This paper is organized as follows: State of the art is a review of the most important works on the subject of automatic bandwidth adjustment and Artificial Intelligence, Materials and Methods discusses the tools used to develop this work, Results and Discussion presents the results of the general and specific experiments that were conducted to determine the feasibility of the Artificial Intelligence methods chosen, finally the Conclusions of this work are exposed.

## II. STATE OF THE ART

There are some authors who study the automatic bandwidth adjustment and integration with artificial intelligence tools, for example, the impact on ISPs is analyzed in [3], where it is identified that the automatic bandwidth adjustment is a technique that emerged due to the significant increase in the amount of traffic on the network permits adjust the bandwidth of a network, depending on the existing traffic demand, allowing access to products and services that with a fixed bandwidth would not be achieved with the same quality of service such as: Streaming, IP voice calls, online games, etc. Shows also that despite the implementation of the automatic adjustment techniques, ISPs have found it necessary to increase physical bandwidth to support the amount of network traffic generated by the applications and services that are booming.

Automatic adjustment ensures that the network will always have adequate bandwidth to work efficiently when users require it, working with adequate bandwidth is important to avoid problems related to network performance such as packet loss or delays [4].

Working with the appropriate bandwidth to meet the demand of each user, allows the reduction of costs related to increased bandwidth consumption without compromising the required quality of service, i.e. it is not necessary to acquire more bandwidth than is truly required by users [5].

The field of Artificial Intelligence focuses its study to develop theories, techniques and methods to mimic and improve human cognitive ability. The rise of Artificial Intelligence tools and technologies provides a more interesting option in the field of computer networks, since it is more efficient to allocate resources using Intelligence algorithms [6].

The integration of Artificial Intelligence and networks is necessary, since the network industry takes giant steps with

each generation, so the combination of these areas permits more efficient, faster networks, with the incorporation of new services and better security. Fields such as wireless networks are influenced by using Artificial Intelligence techniques, which make it possible to exploit multiple opportunities [7], as shown in Fig 1.

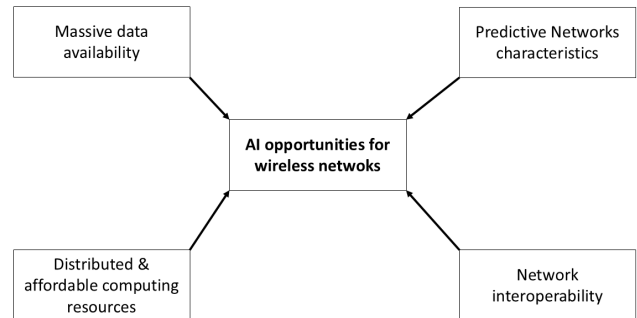


Fig. 1. AI Opportunities in Wireless Networks

Multiple scientific articles have been published referring to the use of Artificial Intelligence methods applied to wireless networks, these mainly analyze how to improve and optimize the performance of networks, in Table 1, shows some methods of Artificial Intelligence, used in the context of wireless networks, for the prediction of various data, ranging from intrusions in the network, resource allocation and network efficiency [8].

TABLE I  
AI METHODS USED IN WIRELESS NETWORKS

Method	Characteristics	Result
Vector Machines Support (SVM)	Construct a high-dimensional feature space from the input vectors in terms of the hyperplane and segregate them into two classes, i.e. Positive and negative	Increased accuracy in the classification of abnormal events in the context of networks
Logistic regression	Predicting the occurrence of an event based on the concept of probability.	Minimize the number of specific features, thus improving logistic regression performance.
KNN algorithm	Sorts new data according to similarity and votes on data points according to euclidean distance measurement	It shows a higher accuracy of 15% in classifying data with varied characteristics.
Nayve Bayes	To classify, calculate the probability to determine the likelihood using bayes' theorem.	It showed good accuracy in the data set with highly variable characteristics.
Random Forest	It is mainly based on multiple decision trees, which is one of the key models in the ml architecture.	It shows a fairly good performance, although the time required is longer than that of other sorters.
Neural Network	It is based on the implementation of multiple nodes, connected to each other, to transmit signals, each piece of information goes through these nodes, where it is subjected to different analyses in order to make a decision.	Achieved an overall read bandwidth improvement of 65.7%.

The implementation of Artificial Intelligence algorithms such as Random Forest and Neural Network in short-range 100/400 Gbit/s data transmission networks has allowed constant and effective monitoring of signal quality, especially in heterogeneous optical networks, effectively identifying the modulation format, achieving an astonishing 100% effectiveness [9].

An algorithm based on Naive Bayes has been successfully implemented in 5G wireless networks. This algorithm divides the network into smaller and more efficient network portions, guaranteeing the adequate optimization of resources, being able to predict the best portion of the network even when there are network interruptions [10].

All the Artificial Intelligence methods analyzed in this work demonstrate a considerable degree of efficiency when working with bandwidth data in scenarios such as resource allocation, network security and resource optimization, taking into consideration the appropriate QoS.

QoS focuses on efficiently allocating resources to ensure optimal performance in a network, since not all Internet traffic is equal, QoS prioritizes different types of traffic depending on their importance when browsing. Failure to apply the appropriate quality of service would result in bad user experience, since there would not be adequate concurrence between the different types of traffic present, so it is important to establish the required QoS parameters [8].

The QoS parameters required for multimedia traffic in a network may vary depending on the context in which they are applied [9], but are generally the following:

**Packet loss:** packets may be dropped when a packet queue overflows, meaning that there are too many packets waiting to be sent, and some must be dropped to avoid further delay in transmission.

**Jitter:** Jitter occurs because of network congestion, variations in transmission times and changes in data paths.

**Latency:** The time it takes for a data packet to travel from its point of origin to its destination in a network.

**Bandwidth:** The capacity of a network communications link to transfer the maximum amount of data from one point to another in a specific time interval.

The present research makes a study of the AI tools to determine which are the most suitable parameters according to the controlled scenario of a 5G wireless network, which has been compiled and is presented in the Dataset.

### III. MATERIALS AND METHODS

The use of the right Dataset is of utmost importance when working on an Artificial Intelligence project, because it influences the training capacity of a model, generalization and performance evaluation.

On the Internet there are millions of Datasets available on various topics, it is practically impossible to search every page where Datasets are available, so the use of a good tool that allows the search of Datasets is important, in this research GoogleDataset and Kaggle are used. Tools that allow searching Datasets, the following key phrases were used: “Bandwidth Dataset for used Artificial Intelligence methods” and “Network computers Dataset for used Artificial Intelligence methods”, resulting in a preliminary result of about 100 Datasets founded.

For the selection of the appropriate Dataset, the parameters were established and the variables identified as shown in Table 2.

The analyzed datasets are the following: “Conference Call Bandwidth Consumption” [10], “International Internet bandwidth per Internet user, kb/s” [11], Dataset 5g Network Metrics High Traffic Event [12]. Of all the Datasets examined, only 1 Dataset met the parameters set out above, this Dataset is called “5g network metrics high traffic event”.

The Dataset has 26 variables where each variable holds 1000 data, giving a total of 26000 data, Table 2 shows the function of each variable selected from the Dataset.

TABLE II  
FUNCTION OF DATASET VARIABLES

FUNCION	VARIABLES
Signal and Quality Indicators	RSRP, RSRQ, CQI
Resource Utilization	BW_Utilization(%), RB_Allocation
Status and User Demand	UE_Demand (kbps)
Traffic and Latency	Traffic_Load (kbps), Latency (ms)
Schemes and Classes	MCS, QoS_Class
Additional Context	Channel_Conditions

Knowing each of the Dataset variables, it is necessary to select the best variables that will be used in the application of experiments to determine the prediction of automatic bandwidth adjustment by means of Artificial Intelligence methods, therefore, using criteria such as: logical discard, knowledge of the subject and correlation analysis, the following variables were chosen:

- Traffic\_Load (kbps): Network traffic load.
- Latency (ms): Network latency, measured in milliseconds (ms).
- BW\_Utilization (%): Percentage of bandwidth utilization.
- CQI (Channel Quality Indicator): Communication channel quality.
- MCS (Modulation and Coding Scheme): Data rate and signal robustness.
- RSRP (Reference Signal Received Power): Signal quality.
- RSRQ (Reference Signal Received Quality): Signal quality with noise and interference.
- RB\_Allocation (Resource Block Allocation): Allocation of resource blocks in the network.
- QoS\_Class (Quality of Service Class): Quality of Service Class assigned to the traffic.
- UE\_Demand (kbps): Bandwidth demand by users.
- Channel\_Conditions: Communication channel conditions.

The analysis of the selected variables shows that there is no variable referring to bandwidth adjustment, which means that the Dataset lacks an objective variable on which, in theory, the necessary predictions should be made to support the research, which is why the decision was made to create such a variable. However, before creating such variable, it should be taken into account that a formula required to calculate the proactive bandwidth adjustment is not a standard formula, since it depends largely on the specific context of the network and the objectives

to be achieved; however, in a 5G network, a weighted combination of different network performance variables can be used to determine the need for a proactive adjustment [13], to calculate the proactive adjustment in this specific network the following variables were taken: Traffic\_Load, Latency and BW\_Utilization, which through a normalization process and their combination by means of the following Python code shown in Figure 2, allowed calculating the proactive adjustment for each case

```
import pandas as pd
import numpy as np
from Orange.data import Table, Domain, ContinuousVariable, DiscreteVariable

# Convertir los datos de Orange a un DataFrame de pandas
df = pd.DataFrame(in_data.X, columns=[var.name for var in
in_data.domain.attributes])

# Si hay columnas objetivo, añadirlas al DataFrame
if in_data.domain.class_vars:
    target_columns = [var.name for var in in_data.domain.class_vars]
    df_target = pd.DataFrame(in_data.Y, columns=target_columns)
    df = pd.concat([df, df_target], axis=1)

# Normalizar las columnas relevantes
df['Traffic_Load (kbps)'] = (df['Traffic_Load (kbps)'] - df['Traffic_Load
(kbps)'].mean()) / df['Traffic_Load (kbps)'].std()
df['Latency (ms)'] = (df['Latency (ms)'] - df['Latency (ms)'].mean()) /
df['Latency (ms)'].std()
df['BW_Utilization (%)'] = (df['BW_Utilization (%)'] - df['BW_Utilization
(%)'].mean()) / df['BW_Utilization (%)'].std()

# Crear la variable objetivo basada en la suma ponderada de los factores
normalizados
df['Proactive_Adjustment'] = (0.5 * df['Traffic_Load (kbps)'] + (0.3 *
df['Latency (ms)'] + (0.2 * df['BW_Utilization (%)']))
```

Fig. 2. Python code, to calculate the proactive adjustment

In the code the variable “Proactive\_Adjustment” stores the amount of bandwidth required at each instant, it is calculated based on latency, traffic load and bandwidth utilization percentage. The variable “Adjustment\_Class” stores the categories assigned to each adjustment range, the adjustment ranges and the categories assigned to each range can be seen in Table 3.

TABLE III  
CATEGORIES ASSIGNED TO EACH ADJUSTMENT RANGE

ADJUSTMENT RANGE	CATEGORY
Less than -0.5	Low
Enter -0.5 and 0	Moderate
Enter 0 and 0.5	High
Greater than 0.5	Critical

#### IV. RESULTS AND DISCUSSION

For the present work two sets of experiments were made with 6 training experiments, to determine which is the best AI method to work with the selected dataset based on a list of 6 previously defined methods: Random Forest, Nayve Bayes, Logistic Regression, SVM, KNN and neural network, where cross validation was used to determine the effectiveness of each method. Experiment 1 was validated by 2 folds, experiment 2 was validated by 5 folds, experiment 3 by 10 folds, experiment 4 by 20 folds. From this point onwards the following

experiments were performed using random data samples, in experiment 5 a sample of 60% was used and in experiment 6 a sample of 70% was used. Table 4 shows the main results of the training experiments.

TABLE IV  
RESULTS OF TRAINING EXPERIMENTS IN ACCURACY

	ACCURACY						
	Random Forest	Nayve Bayes	Logistic Regression	SVM	K-NN	Neural Network	
EXPERIMENTS	2 folds	95.8%	95.0%	90.7%	84.5%	22.5%	28.0%
	5 folds	99.1%	95.0%	93.2%	88.8%	22.9%	28.0%
	10 folds	99.2%	95.1%	92.5%	88.3%	22.4%	28.0%
	20 folds	99.0%	94.8%	92.8%	89.4%	22.2%	28.0%
	60 % data random	99.2%	95.7%	91.3%	88.9%	24.4%	28.0%
	70 % data random	98.8%	94.7%	92.0%	84.9%	24.1%	28.0%

In the 6 validation experiments that serve to validate the results of the general experiments, by applying each AI method separately to the dataset, then in the specific experiment 1 the Random Forest method was used, in the specific experiment 2 the Nayve Bayes method was used, in the specific experiment 3 the Logistic Regression method was used, in the specific experiment 4 the SVM method was used, in the specific experiment 5 the KNN method was used and in the specific experiment 6 the Neural Network method was used, in the Table 5 shows the main results of the validation experiments.

TABLE V  
RESULTS OF VALIDATION EXPERIMENTS

MODEL	TRAINING: 600 DATES		VALIDATION: 400 DATES	
	Successes	Mistakes	Successes	Mistakes
Random Forest	600 (100%)	0 (0%)	391 (97.8%)	9 (2.2%)
Nayve Bayes	560 (93.3%)	40 (6.7%)	379 (94.8%)	21 (5.2%)
Logistic Regression	567 (94.5%)	33 (5.5%)	368 (92%)	32 (8%)
SVM	587 (97.8%)	13 (2.2%)	349 (87.3%)	51 (12.7%)
KNN	288 (48%)	312 (52%)	89 (22.3%)	311(77.7%)
Red Neuronal	169 (28.1%)	431(71.9%)	111 (27.8%)	289 (72.2%)

#### A. Training Experiments

The Orange Data Mining environment [14] was used to perform the experiments, firstly, the Testing option is used to test the effectiveness of various prediction models when working with the data stored in the Dataset, then several Artificial Intelligence techniques must be chosen to perform the respective testing, the techniques chosen are: neural network, logistic regression, Random Forest, Support Vector Machines (SVM), Nayve Bayes and k-nearest neighbor (KNN) as shown in Fig. 3. To determine the effectiveness of each model, several evaluation metrics proposed by [15] are used.

Precision (Prec): This metric focuses on the positive and false positive values; it is obtained from the total true positives divided by the sum of true positives and false positives (Eq. 1).

$$Precision = \frac{True\ positives}{True\ positives + False\ Positives} \tag{1}$$

Recall: With the same principle as precision, but with the difference that it focuses on false negatives. It is calculated from the total of true positives divided by the sum of true positives and false negatives (Eq. 2).

$$Recall = \frac{True\ positives}{True\ positives + False\ Negatives} \tag{2}$$

F1 Score (F1): This metric takes into consideration precision and recall. It is obtained from the double product of the multiplication of precision and recall divided by the sum of precision and recall (Eq. 3).

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3}$$

Accuracy (CA): Refers to the number of correct predictions divided by the number of total predictions (Eq. 4).

$$Accuracy = \frac{Correct\ predictions}{Total\ Predictions} \tag{4}$$

Area under the ROC curve (AUC): This metric represents the probability that a randomly chosen positive-valued sample has a higher rating by the model than a randomly chosen negative-valued sample. A perfect model would have an AUC = 1 [16].

Matthews Correlation Coefficient (MCC): It is a statistical metric, widely used in binary classifications, its value ranges from -1 to +1, it returns good results only if the 4 values of the confusion matrix are also good [17].

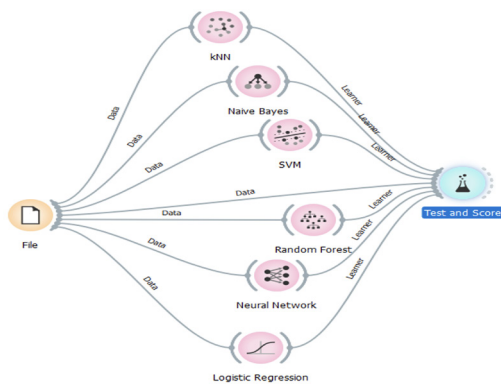


Fig. 3. Testing of Various AI Methods

### 1. Experiment 1

For the present experiment the cross validation was used, with a number of folds of 2, giving as a result what is shown in Table 6, where the models are placed according to the accuracy they showed in the testing, where it can be clearly observed the superiority of the Random Fo-

rest model, over the other models, giving a much better performance than the others, while models such as Naive Bayer, Logistic Regression and SVM show a fairly good reliability, on the contrary, the KNN and Neural Network models show very poor results with respect to the other models.

TABLE VI  
RESULTS OF EXPERIMENT 1

MODEL	AUC	CA	F1	PREC	RECALL	MCC
Random Forest	0.999	0.985	0.985	0.985	0.985	0.980
Nayve Bayes	0.988	0.950	0.950	0.952	0.950	0.934
Logistic Regression	0.991	0.907	0.907	0.907	0.907	0.876
SVM	0.979	0.854	0.855	0.856	0.854	0.805
KNN	0.477	0.225	0.220	0.227	0.225	-0.037
Neural Network	0.514	0.280	0.123	0.078	0.280	0.000

### 2. Experiment 2

For the second experiment, cross-validation was used, but this time the number of folds was increased to 5, resulting in the results shown in Table 7, where a clear increase in accuracy can be observed with respect to the previous experiment, leading again the Random Forest model, which remains as the model with the highest accuracy in all aspects, followed by the Naive Bayes and Logistic Regression models, which remain with a good reliability, the SVM model has significantly increased its reliability, while the KNN and neural network models remain as the model with the lowest accuracy.

TABLE VII  
RESULTS OF EXPERIMENT 2

MODEL	AUC	CA	F1	PREC	RECALL	MCC
Random Forest	1.000	0.991	0.991	0.991	0.991	0.988
Nayve Bayes	0.989	0.950	0.950	0.953	0.950	0.934
Logistic Regression	0.994	0.932	0.932	0.932	0.932	0.909
SVM	0.986	0.888	0.888	0.889	0.888	0.850
KNN	0.489	0.229	0.227	0.233	0.229	-0.030
Neural Network	0.518	0.280	0.123	0.078	0.280	0.000

### 3. Experiment 3

For the third experiment the cross validation was used, but this time the number of folds was increased to 10, again there is an increase in the accuracy of the Random Forest model remaining as the model with the highest reliability, followed by the Naive Bayes, Logistic Regression and SVM models with a very good reliability, while the KNN model and the neural network remain as the models with the lowest reliability, this can be seen in Table 8.

TABLE VIII  
RESULTS OF EXPERIMENT 3

MODEL	AUC	CA	F1	PREC	RECALL	MCC
Random Forest	1.000	0.992	0.992	0.992	0.992	0.989
Nayve Bayes	0.989	0.951	0.951	0.954	0.951	0.936
Logistic Regression	0.994	0.925	0.925	0.925	0.925	0.900
SVM	0.986	0.883	0.883	0.884	0.883	0.844
KNN	0.493	0.224	0.221	0.227	0.224	-0.037
Neural Network	0.506	0.280	0.123	0.078	0.280	0.000

#### 4. Experiment 4

For the fourth experiment we used the cross validation which is automatically executed by the Orange Data Mining environment, this time we increased the number of folds to 20, the trend of the previous experiments remains almost the same, but unlike the previous experiments, we noticed a slight decrease in the reliability of the Random Forest, Nayve Bayes and KNN models, a slight decrease in the reliability of the Random Forest, Nayve Bayes and KNN models is noticed, on the other hand the reliability of the Logistic Regression and SVM models has slightly increased, while the neural network remains the same being the model with the lowest reliability, this can be observed in [Table 9](#).

TABLE IX  
RESULTS OF EXPERIMENT 4

MODEL	AUC	CA	F1	PREC	RECALL	MCC
Random Forest	1.000	0.990	0.990	0.990	0.990	0.987
Nayve Bayes	0.989	0.948	0.948	0.951	0.948	0.932
Logistic Regression	0.994	0.928	0.928	0.928	0.928	0.904
SVM	0.985	0.894	0.894	0.895	0.894	0.858
KNN	0.488	0.222	0.220	0.226	0.222	-0.039
Neural Network	0.503	0.280	0.123	0.078	0.280	0.000

#### 5. Experiment 5

By using a random data sample of 60% of the Dataset, with 10 test repetitions, we obtained the results shown in [Table 10](#), where maintaining the trend marked in the previous experiments, we note the clear superiority in reliability of the Random Forest model, which shows much higher metrics than the other models, on the other hand models such as Nayve Bayes, Logistic Regression and SVM maintain a very similar level to those shown in the experiments and the KNN and neural network models remain as the models with the lowest reliability.

TABLE X  
RESULTS OF EXPERIMENT 5

MODEL	AUC	CA	F1	PREC	RECALL	MCC
Random Forest	1.000	0.992	0.992	0.992	0.992	0.989
Nayve Bayes	0.990	0.957	0.956	0.959	0.957	0.943
Logistic Regression	0.991	0.913	0.913	0.913	0.913	0.884
SVM	0.985	0.889	0.889	0.891	0.889	0.852
KNN	0.490	0.244	0.242	0.250	0.244	-0.009
Neural Network	0.517	0.280	0.123	0.078	0.280	0.000

#### 6. Experiment 6

By using a random data sample of 70% of the Dataset, with 10 test repetitions, we obtained the results shown in [Table 11](#), where we can notice a small decrease in the reliability of most prediction models, except for the Logistic Regression model and SVM, which have experienced a small increase in their reliability, while the neural network remains the same. Data charts which are typically black and white but sometimes include color.

TABLE XI  
RESULTS OF EXPERIMENT 6

MODEL	AUC	CA	F1	PREC	RECALL	MCC
Random Forest	1.000	0.988	0.988	0.988	0.988	0.984
Nayve Bayes	0.987	0.947	0.947	0.951	0.947	0.931
Logistic Regression	0.993	0.920	0.920	0.920	0.920	0.893
SVM	0.986	0.894	0.894	0.895	0.894	0.858
KNN	0.495	0.241	0.239	0.248	0.241	-0.014
Neural Network	0.509	0.280	0.123	0.078	0.280	0.000

#### B. Validation Experiments

Having known the effectiveness of various models through initial testing to see the behavior of the data against each model, it is necessary to validate these results, in order to give credibility to the results given by the general testing.

An abbreviation will be assigned to each word in the tables for better presentation.

- Cat (Category)
- Suma (Summation)
- Hi (High)
- Lo (Low)
- Cri (Critical)
- Mo (Moderate)

#### 1. Random Forest

It is a technique that consists of the creation of multiple decision trees to be trained on a different set of data, and then all the results are integrated to give a final answer. For the present evaluation, the Random Forest model was used on a sample of 60% of the data for training, giving the results shown in [Table 12](#).

TABLE XII  
TRAINING RESULTS BY RANDOM FOREST METHOD

	CAT	PREDICTION				SUMA
		HI	LO	CRI	MO	
CURRENT STATUS	HI	164	0	0	0	164
	LO	0	121	0	0	121
	CRI	0	0	146	0	146
	MO	0	0	0	169	169
	SUMA	164	121	146	169	600

Now that the respective model training has been performed, it is necessary to validate the results by evaluating the model by applying 40% of the remaining data, in order to check its efficiency, these new results can be seen in Table 13.

TABLE XIII  
EVALUATION RESULTS USING RANDOM FOREST METHOD

	CAT	PREDICTION				SUMA
		HI	LO	CRI	MO	
CURRENT STATUS	HI	94	0	0	2	96
	LO	0	102	0	4	106
	CRI	2	0	85	0	87
	MO	1	0	0	110	111
	SUMA	97	102	85	116	400

The results of the application of the Random Forest technique for the automatic bandwidth adjustment prediction model show that, although in the initial training it was able to get 100% of the predictions right, at the time of validation it made few errors in the predictions.

### 2. Nayve Bayes

It is a technique that uses Bayes' theorem to determine the most probable membership of a class. For the present training, the Nayve Bayes technique was used, together with a sample of 60% of the data, achieving the results shown in Table 14.

TABLE XIV  
TRAINING RESULTS USING THE NAYVE BAYES METHOD

	CAT	PREDICTION				SUMA
		HI	LO	CRI	MO	
CURRENT STATUS	HI	150	0	4	10	164
	LO	0	121	0	0	121
	CRI	0	0	146	0	146
	MO	0	26	0	143	169
	SUMA	150	147	150	153	600

Now the respective validation must be performed, by means of a model evaluation, applying 40% of the remaining data to check the efficiency of the model, when trained, the results of the validation can be seen in Table 15.

TABLE XV  
EVALUATION RESULTS USING THE NAYVE BAYES METHOD

	CAT	PREDICTION				SUMA
		HI	LO	CRI	MO	
CURRENT STATUS	HI	87	0	1	8	96
	LO	0	106	0	0	106
	CRI	0	0	87	0	87
	MO	0	12	0	99	111
	SUMA	87	118	88	107	400

The results of the Nayve Bayes technique, after the respective validation, show that, both in the training and in the evaluation, the model made several prediction errors, and it can be deduced that the reliability is quite high because it made few errors.

### 3. Logistic Regression

It is a technique that, through the use of mathematics, can find relationships between data factors, and then predict one value based on the other value. To train the model based on this technique, a sample of 60% of the data was used, obtaining the results shown in Table 16.

TABLE XVI  
TRAINING RESULTS USING LOGISTIC REGRESSION METHOD

	CAT	PREDICTION				SUMA
		HI	LO	CRI	MO	
CURRENT STATUS	HI	149	0	10	5	164
	LO	0	117	0	4	121
	CRI	5	0	141	0	146
	MO	5	4	0	160	169
	SUMA	159	121	151	169	600

As can be seen in Table 16, the prediction model based on the Logistic Regression technique, has made many more errors in the predictions, now it is necessary to make the testing process by using the remaining 40% of data, in Table 17 you can see the results of the test but sometimes include color.

TABLE XVII  
EVALUATION RESULTS USING LOGISTIC REGRESSION METHOD

	CAT	PREDICTION				SUMA
		HI	LO	CRI	MO	
CURRENT STATUS	HI	89	0	4	3	96
	LO	0	92	0	14	106
	CRI	5	0	82	0	87
	MO	4	2	0	105	111
	SUMA	98	94	86	122	400

Through the testing process, it has been proven that the model has made several errors in the predictions. It is noteworthy that the model has made more errors in the moderate classification, since it places 17 data wrongly,

which belong to other categories, but the number of correct predictions is much higher than the errors, therefore, the reliability of the model is high.

#### 4. Support Vectorial Machine SVM

It is a technique that helps to predict outliers in different groups, always looking for the best plane to separate these groups in a space of many qualities. For the training of the model based on the SVM technique, a sample of 60% of the available data was used, obtaining the results presented in Table 18.

TABLE XVIII  
TRAINING RESULTS USING THE SVM METHOD

CURRENT STATUS	CAT	PREDICTION				SUMA
		HI	LO	CRI	MO	
CURRENT STATUS	HI	160	0	1	3	164
	LO	0	119	0	2	121
	CRI	4	0	142	0	146
	MO	3	0	0	166	169
SUMA		167	119	143	171	600

Having performed the training process of the SVM-based model for the prediction of the automatic bandwidth adjustment, it is necessary to perform the validation process by applying the respective test using 40% of the remaining data, obtaining the results presented in Table 19.

TABLE XIX  
EVALUATION RESULTS USING SVM METHOD

CURRENT STATUS	CAT	PREDICTION				SUMA
		HI	LO	CRI	MO	
CURRENT STATUS	HI	86	0	6	4	96
	LO	1	91	0	14	106
	CRI	8	0	79	0	87
	MO	14	4	0	93	111
SUMA		109	95	85	111	400

The results of the evaluation indicate that most of the model's errors are concentrated in the assignment of data to the "High" category, while 23 of these data belong to other categories; therefore, it can be said that the reliability of this model is medium-high.

#### 5. K- Nearest Neighbor (KNN)

It is a technique based on the implementation of an algorithm to make predictions by searching for similar data learned in the training stage. To train a model based on this technique, a sample of 60% of the available data was used, obtaining the results presented in Table 20.

TABLE XX  
RESULTS OF TRAINING USING KNN METHOD

CURRENT STATUS	CAT	PREDICTION				SUMA
		HI	LO	CRI	MO	
CURRENT STATUS	HI	106	12	29	17	164
	LO	39	50	14	18	121
	CRI	43	24	64	15	146
	MO	48	17	36	68	169
SUMA		236	103	143	118	600

In the present model, it is observed that the training results are not as expected, since there are many prediction errors, but the corresponding validation should be done by applying a test with 40% of the remaining data, obtaining the results shown in Table 21.

TABLE XXI  
EVALUATION RESULTS USING THE KNN METHOD

CURRENT STATUS	CAT	PREDICTION				SUMA
		HI	LO	CRI	MO	
CURRENT STATUS	HI	28	18	24	26	96
	LO	44	20	25	17	106
	CRI	41	15	17	14	87
	MO	35	25	27	24	111
SUMA		148	78	93	81	400

The results are quite clear, the reliability of this model is very low, since the errors made in each prediction exceed the hits, so it can be deduced that the KNN algorithm is very ineffective for working with this type of data.

#### 6. Neural Network

This technique consists of a set of nodes connected to each other to transmit signals; each piece of information passes through these nodes, where it is subjected to different analyses in order to make a decision. For the training of an automatic bandwidth adjustment prediction model based on a neural network, a sample equivalent to 60% of the available data was used, obtaining the results shown in Table 22.

TABLE XXII  
TRAINING RESULTS USING NEURAL NETWORK METHOD

CURRENT STATUS	CAT	PREDICTION				SUMA
		HI	LO	CRI	MO	
CURRENT STATUS	HI	0	0	0	164	164
	LO	0	0	0	121	121
	CRI	0	0	0	146	146
	MO	0	0	0	169	169
SUMA		0	0	0	600	600



The training results are very unsatisfactory, since the neural network achieved very few predictions correctly, since all the predictions were assigned in the “moderate” category, but only 169 out of 600 are assigned correctly, in this sense, the corresponding validation must be performed by using 40% of the remaining data, Table 23 shows the results of the validation process.

TABLE XXIII  
EVALUATION RESULTS USING NEURAL NETWORK METHOD

		PREDICTION				
		CAT	HI	LO	CRI	MO
CURRENT STATUS	HI	0	0	0	96	96
	LO	0	0	0	106	106
	CRI	0	0	0	87	87
	MO	0	0	0	111	111
	SUMA	0	0	0	400	400

The validation results confirm the training results, in the new validation predictions the model commits the same errors of the training phase, classifying all the data in the “Moderate” category, of which 111 are correctly assigned out of a total of 400, therefore the reliability of the model is low.

C. Discussion

Once all the general and specific experiments are completed, the results of the experiments are analyzed in each of the tables presented above. It can be observed that the result of the experiments varies very little in each of the AI models used, with the Random Forest, Naive Bayes and Logistic Regression models standing out after having achieved predictions greater than 92% effectiveness in the general and specific experiments, while the SVM model demonstrated good effectiveness reaching 88.8% of correct predictions in the general experiments, varying greatly at the time of performing the specific experiment, rising to 97.8% at the time of training and then having a reduction to 87.3% when completely new data was presented. On the other hand, the KNN and neural network models show an effectiveness of less than 50% of correct predictions in all experiments, all the above can be seen in Table 24.

TABLE XXIV  
COMPARISON OF TRAINING AND VALIDATION RESULTS

MODELS		ACCURACY		
		5 folds	Training: 600 Dates	Validation: 400 dates
	Random Forest	99.1%	100%	97.8%
	Nayve Bayes	95.0%	93.3%	94.8%
	Logistic Regression	93.2%	94.5%	92%
	SVM	88.8%	97.8%	87.3%
	KNN	22.9%	48%	22.3%
	Neural Network	28.0%	28.1%	27.8%

Based on the results Random Forest, Naive Bayes, Logistic Regression and SVM models tend to be more efficient in predicting the automatic bandwidth adjustment since they balance well the complexity of the model with the ability to generalize and handle noisy data, on the other hand, KNN and neural network can become inefficient due to the high complexity and computational overhead that is not always necessary for this type of problem.

V. CONCLUSION

Upon careful analysis of the Dataset used for the experiments, it was determined that it partially complies with the required metrics, since it lacks the variable required for automatic adjustment, being necessary to create such variable and assign the adjustment categories to be used as a target variable.

Through the application of the general experiments, it was observed that: Random Forest, Nayve Bayes, Logistic Regression and SVM, demonstrated an effectiveness above 90% measured through the parameters: precision (PREC), classification accuracy (CA), area under the ROC curve (AUC), RECALL and Mathematical Correlation Coefficient (MCC), being that Random Forest demonstrated an effectiveness of 99% in all parameters, being the highest with respect to the other methods, while KNN and Neural Network demonstrated effectiveness below 50%.

The specific experiments showed that the most effective method for the prediction of automatic bandwidth adjustment is Random Forest, since in the training it approached 100% of the predictions, while in the evaluation it got 97.8% of the predictions right, with higher percentages of correct predictions in training and evaluation compared to the other methods,

Comparing the different Artificial Intelligence techniques, taking into consideration the quality of service for the automatic or dynamic bandwidth adjustment, exposes the limitations and strengths that each of the evaluated techniques have when working with certain types of data, although some techniques show a very high prediction effectiveness regardless of the training and validation data, on the contrary, other techniques show that there are variations when exposing the model to training and validation data, being that there is a notable difference between the results of training and validation.

VI. FUTURE WORKS

Future work includes developing an automatic bandwidth adjustment model based on Artificial Intelligence methods, focusing in Analysis of relationship the Random Forest and Neural Network models and the multimedia traffic specially for wireless networks.

REFERENCES

[1] T. Alalibo, O. Sunny and E. Promise, “Bandwidth optimization of wireless networks using,” March 2020. [Online]. Available: <https://www.irejournals.com/formatedpaper/1702018.pdf>

[2] S. C. Madanapalli, H. H. Gharakhieli and V. Sivaraman, “Inferring Netflix user experience from broadband network measurement,” 5 August 2019. [Online]. <https://doi.org/10.23919/TMA.2019.8784609>

- [3] S. Rai and A. Kumar, "Dynamic bandwidth allocation in optical networks using Machine Learning," July 2021. [Online]. <https://doi.org/10.17605/OSF.IO/5UKBM>
- [4] A. A. Suham and H. H. Karar, "Optimization of Resource and Bandwidth Allocation in Wireless Networks Performance Analysis using Artificial Intelligence," 2020. [Online]. Available: [https://repository.atu.edu.iq/uploads/repo\\_file\\_7\\_23\\_10\\_21\\_51.pdf](https://repository.atu.edu.iq/uploads/repo_file_7_23_10_21_51.pdf).
- [5] J. Cao, Z. Ma, J. Xie, X. Zhu, F. Dong and Boliu, "Towards Tenant Demand-Aware Bandwidth Allocation Strategy in Cloud Data Center," April 2020. [Online]. Available: <https://www.sciencedirect.com/science/article/abs/pii/S0167739X17311834>.
- [6] K. Lin, C. Li, D. Tian, A. Ghoneim, M. S. Hossain and S. O. Amin, "Artificial-Intelligence-Based Data analytics for cognitive communication in heterogeneous wireless networks," June 2019. [Online]. <https://doi.org/10.1109/MWC.2019.1800351>
- [7] B. L. Khaled, C. Wei, S. Yuanming, Z. Jun and A. Z. Ying-Jun, "The Roadmap to 6G – AI Empowered," 19 July 2019. [Online]. HYPERLINK <https://doi.org/10.48550/arXiv.1904.11686>; <https://doi.org/10.48550/arXiv.1904.11686>
- [8] T. Sowmya and A. Mary, "A comprehensive review of AI based intrusion detection system," agosto 2023. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2665917423001630#sec3>.
- [9] J. Zhao, Z. Su, Y. Yang and T. Xu, "ANN/Random forest based performance monitoring in high-speed short-reach optical interconnections," October 2024. [Online]. HYPERLINK "<https://doi.org/10.1016/j.yofte.2024.103941>" \t "\_blank" \o "Persistent link using digital object identifier" <https://doi.org/10.1016/j.yofte.2024.103941>
- [10] Wezen, "QoS, qué es y cuál es su importancia en la gestión IT," 15 August 2023. [Online]. Available: <https://www.wezengroup.com/qos-cual-es-su-importancia-en-la-gestion-it/>.
- [11] VASExperts, "Cómo utilizar QoS para garantizar la calidad del acceso a Internet," 23 January 2023. [Online]. Available: <https://vasexperts.com/es/blog/quality-of-service/how-to-use-qos-to-ensure-the-internet-access-quality/>.
- [12] D. Young, "Conference Call Bandwidth Consumption. [Dataset]," (2020). [Online]. Available: <https://www.kaggle.com/dikamsiyoung/conference-call-bandwidth-consumption/data>.
- [13] Union International Telecommunication, "International Internet bandwidth per Internet user, kb/s. [Dataset]," 2016. [Online]. Available: [https://www.theglobaleconomy.com/Slovenia/Internet\\_bandwidth/](https://www.theglobaleconomy.com/Slovenia/Internet_bandwidth/).
- [14] K. Raghunath, "5g network metrics high traffic event. [Dataset]," 18 August 2023. [Online]. <https://doi.org/10.21227/1ryt-wb82>
- [15] L. Mei, J. Gou, Y. Cai, H. Cao and L. Yong, "Realtime Mobile Bandwidth and Handoff Predictions in 4G/5G Networks," 27 April 2021. [Online]. Available: <https://doi.org/10.48550/arXiv.2104.12959>
- [16] O. Schmitz, "Orange Data Mining: Aprendizaje automático de código abierto y visualización de datos," 9 May 2023. [Online]. Available: <https://www.linkedin.com/pulse/orange-data-mining-aprendizaje-autom%C3%A1tico-de-c%C3%B3digo-y-schmitz/>.
- [17] M. Amer, "Métricas de evaluación de clasificación: precisión, exactitud, recuperación y F1 explicadas visualmente," 7 June 2022. [Online]. Available: <https://cohere.com/blog/classification-eval-metrics>.
- [18] Deepchecks, "Comprensión de las métricas de puntuación F1, precisión, ROC-AUC y PR-AUC para modelos," 13 June 2024. [Online]. Available: <https://deepchecks.com/f1-score-accuracy-roc-auc-and-pr-auc-metrics-for-models/>.
- [19] M. Tech, "¿Qué es el coeficiente de correlación de Matthews (MCC)?," 13 December 2022. [Online]. Available: [https://medium.com/@CuttiE\\_MarU/what-is-matthews-correlation-coefficient-mcc-bb07a94162ba](https://medium.com/@CuttiE_MarU/what-is-matthews-correlation-coefficient-mcc-bb07a94162ba).