Classification of congestion problems in a telecommunications network using its performance indices

Borel SONNA Laure OBAYA Lauren MOUKODI André NYEMB

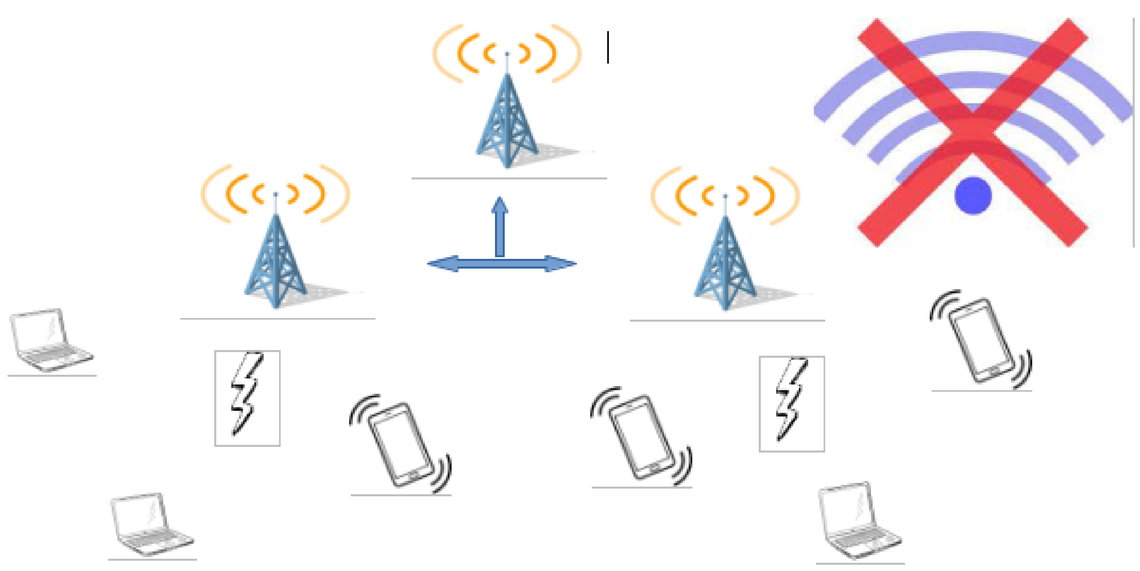
BEL'S AI INITIATIVE Deep Learning Indaba 2023

We are BEL'S AI INITIATIVE

BEL'S AI INITIATIVE for BE EDUCATED and LEARN SKILLS in ARTIFICIAL INTELLIGENCE INITIATIVE) is an association based in Cameroon dedicated to demystifying artificial intelligence in Africa. Our primary objective is to provide specialized content and training in artificial intelligence while initiating various projects aimed at promoting this discipline on the African continent.

Figure 1. Website link

UNDERSTANDING THE MACHINE LEARNING PROBLEM IN THIS PROJECT



- **Congestion** in a telecommunications network occurs when the amount of traffic or data in circulation exceeds the maximum capacity of the network to handle it efficiently. This leads to **degraded performance, transmission delays and disruptions to data delivery, affecting the quality of service offered to users**. In short, congestion occurs when the network is overloaded and cannot handle the volume of data flowing through it.
- That said, congestion problems depend on **the performance indices of a network at a given time**. So the data we will use to train our machine learning model on how to classify congestion problems in the network will be **a big dataset of network's perfomance indices when there is a congestion problem.**
- Our dataset will be a set of **data labelled** for each record with the associated type of congestion. This is **a supervised learning machine learning algorithm.**
- Since the type of congestion is **a qualitative data**, we're going to use a **supervised learning classification algorithm.**

In short, this is a supervised machine learning algorithm for multiclass classification of congestion problems in a telecommunications network using network performance indices.

DESCRIPTION OF THE DATASET(1)

This dataset has been downloaded from the github platform via the following link https://github.com/CallMeAmartya/IITKgp-Interhall-Data-Analytics Its features are as follows:

- In our dataset, we have 39 attributes or columns and 78560 rows or records.
- In our dataset we have 39 attributes, 37 of which are integers (these values are numeric) and the other two are objects that need to be numbered.
- No missing values

We have a fairly clean dataset with no missing values and with almost all the attributes corresponding to the performance indices already in numerical form. It should be noted that only the prediction variable, i.e. the type of congestion, is not numeric, nor is the variable corresponding to the equipment manufacturer who designed the telecommunication equipment on which we have collected the data.

DESCRIPTION OF THE DATASET(2)

This graph shows the distribution of the data in relation to the prediction variable representing the type of congestion.

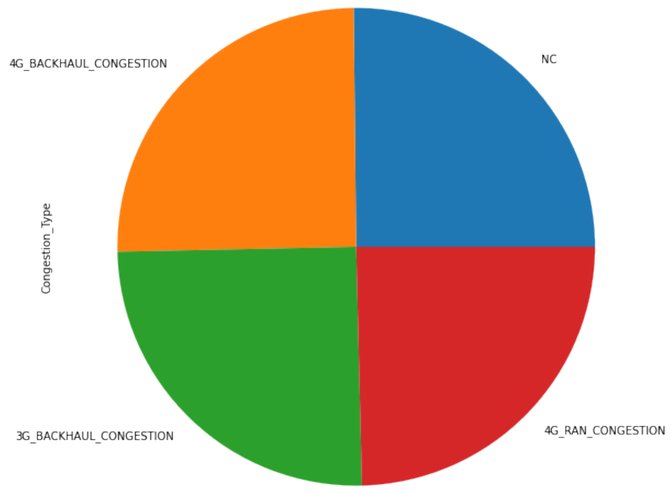


Figure 2. Data repartition

There are 4 classes in our dataset. The dataset is not balanced in terms of the number of records per class. These are the different classes in ascending order of the number of records contained in the dataset:

- 4G-RAN-CONGESTION 19336 records
- 3G-BACKHAUL-CONGESTION 19688 records
- 4G-BACKHAUL-CONGESTION 19765 records
- NC 19771 records

DATA PRE-PROCESSING AND LEARNING PHASE

- **Data discretization:** we match the values of the two attributes with data of type caratere to numbers around 0.
- **Data reduction:** We remove all attributes that are not performance indices for predicting the type of congestion in a network.
- **Data segmentation:** We separate the output variable (type of congestion) from all the other variables which are the input data. We divide the input data evenly by class so as to have a batch of 80% for the training phase and 20% for the test phase.
- **Model selection:** we compare 4 models for this classification during the training phase. These are MLPClassifier, Logistic Regression, Gaussian Naive Bayes and Neural Network. The model selected is the logistic regression because of many informations given by its learning curve and for its best performance.
- **Training phase with logistic Regression**

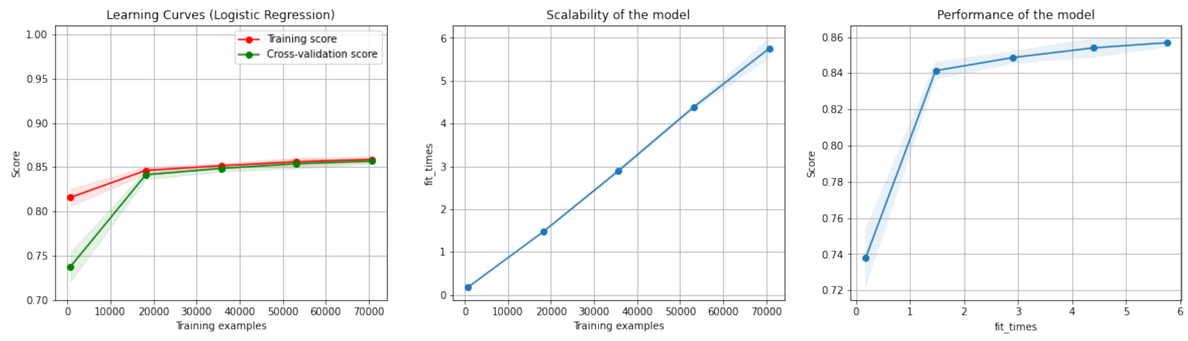


Figure 3. learning curve

TEST PHASE AND PERFORMANCE

The performance of the model is given by this table.

INTERPRETATION DES RESULTATS DE TEST

ON A:

Accuracy:84.76 Classification report precision recall f1-score support

-1 => "NC"	1.00	0.98	0.99	3954
0 => "4G_BACKHAUL_CONGESTION"	0.98	0.72	0.83	3953
1 => "3G_BACKHAUL_CONGESTION"	0.97	0.69	0.81	3938
2 => "4G_RAN_CONGESTION"	0.63	0.99	0.77	3867
accuracy		0.85	15712	
macro avg	0.89	0.85	0.85	15712
weighted avg	0.90	0.85	0.85	15712

Figure 4. performance of the model

CONCLUSION

In conclusion, **BEL'S AI INTIATIVE** is an association committed to promoting artificial intelligence in Africa. At this **Deep Learning Indaba 2023** conference, we have decided to present one of our projects we worked on with our students. The project involves the design and implementation of a machine learning model that classifies the congestion experienced by telecommunications equipment in the network, using performance indices. We obtain a score of abbout 90%. It thus offers the possibility of ensuring predictive and preventive maintenance of telecommunications equipment for network operators.