

MODELING A 3G MOBILE PHONE BASE RADIO USING ARTIFICIAL INTELLIGENCE TECHNIQUES

Eduardo Calo^{1,5}, Gabriel Vaca¹, Cristina Sánchez²,
David Jines³, Giovanni Amancha³, Ángel Flores³,
Alex Santana G³ and Fernanda Oñate⁴

¹Carrera Electricidad, Instituto Superior María Natalia Vaca, Ambato, Ecuador

²Carrera Mecánica Automotriz,

Instituto Superior María Natalia Vaca, Ambato, Ecuador

³Carrera Electrónica, Instituto Superior María Natalia Vaca, Ambato, Ecuador

⁴Instituto Superior Tecnológico Pelileo-Baños, Baños, Ecuador

⁵UNIR, Logroño, La Rioja, España

ABSTRACT

The main objective of this work is to be able to use artificial intelligence techniques to be able to design a predictive model of the performance of a third-generation mobile phone base radio, using the analysis of KPIs obtained in a statistical data set of the daily behaviour of an RBS. For the realization of these models, various techniques such as Decision Trees, Neural Networks and Random Forest were used. which will allow faster progress in the deep analysis of large amounts of data statistics and get better results. In this part of the work, data was obtained from the behaviour of a third-party mobile phone base radio generation of the Claro operator in Ecuador, it should be noted that. The data are KPIs of the daily and hourly performance of a radio base of third generation mobile telephony, these data were obtained through the operator's remote monitoring and management tool Sure call PRS. To specify this practical case, several models were generated based on in various artificial intelligence technique for the prediction of performance results of a mobile phone base radio of third generation, the same ones that after several tests were creation of a predictive model that determines the performance of a mobile phone base radio. As a conclusion of this work, it was determined that the development of a predictive model based on artificial intelligence techniques is very useful for the analysis of large amounts of data in order to find or predict complex results, more quickly and trustworthy.

KEYWORDS

Neural Networks, Performance, Radio Base, Random Forest, Throughput.

1. INTRODUCTION

Deep learning is one of the techniques more evolution has had within the field of Intelligence Artificial, thanks to this advance a development has begun to very notorious in various sectors such as financial, vehicle, telecommunications, understanding of natural language, translators of languages and many more sectors worldwide, streamlining processes optimization and having a future vision of them, optimizing resources.

The purpose of this work is to use these intelligence techniques artificial in a practical case of real life, such as the analysis of the performance of an RBS and predict its behaviour or results to be obtained after the analysis of its operating parameters, the same ones that in the telecommunications area will help us in the optimization of resources and in the improvement of a telephone network third generation, the analysis of this performance was carried out at the level of data traffic in telephone stations, predicting its behaviour with the change of different variables.

To specify this practical case, several models were generated based on in various artificial intelligence techniques such as Random Forest, Decision Trees and Neural Networks, for the prediction of performance results of a mobile phone base radio of third generation, the same ones that after several tests were creation of a predictive model that determines the performance of a mobile phone base radio.

1.1. In the investigation “Intelligent Techniques in the assignment of dynamic spectrum for cognitive wireless networks”:

In this work, a number of techniques of artificial intelligence that help provide solutions in the management of electromagnetic spectrum in wireless networks, taking into account that mobile telephony is within wireless networks due to the use and management of frequencies found in the spectrum, these investigated techniques allow to analyse parameters such as spectrum availability, energy consumption, channel division, necessary requirements for the user, taking into account that all These parameters have as a final result the existence of a transmission and reception wireless communication system optimal data. [1].

1.2. In the research “Cellular mobile telephony: origin, evolution, perspectives”

Worldwide, the evolution in mobile telephony has had great events, such is the case that the use of mobile devices has potentially increased use of mobile phones initially relied on him sending voice signals and text messages in what corresponds to the first generation of mobile telephony, from there onwards there have been various types of evolutions that range from the first generation to the fifth generation, from the third generation the development is very significant that the itself allows the use of mobile phones not only for transmission determined which is the model that would give us the best results when working With this type of data, the variables with which we worked, have relation to data traffic measurements within radio bases, both for data upload and download traffic.

2. STATE OF THE ART

As part of the techniques proposed for the development of this the work evaluates the use of Artificial Intelligence as a means of function of receiving the signals in an RBS both for frequencies of rise and fall in the two types of technologies. [2]

2.1. In the research "Third generation cellular networks CDMA2000 and WCDMA”:

After the emergence of second-generation technology and due to the increase in users and therefore saturation of the network and need to implement new services in what has to do with mobile telephony, third generation cellular networks appear with its CDMA 2000 and WCDMA models, the first of these models is the link to second generation and was in charge of the link of

compatibility and migration between second and third generation, while WCDMA focuses on increasing the number of users, as the picture shows [3]

In CDMA2000 there are two modulation alternatives for the downlink and these are: multicarrier modulation (MC) and direct spread (DS) that use 3 carriers of 1.25 MHz or 5 MHz; while in WCDMA it uses a single 5 MHz carrier.[4]

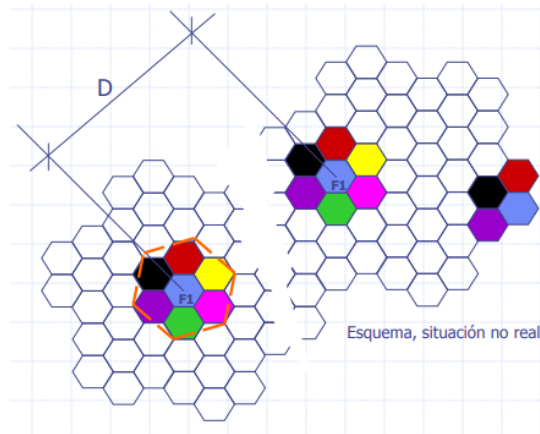


Figure 1. Third generation cellular networks CDMA2000 and WCDMA [3]

2.2. In the research “Intelligent Systems applied to Data”

The emergence of artificial intelligence and the use of all its techniques and tools have resulted in the emergence of intelligent systems the same ones that are capable of learning and execute actions automatically based solely on examples of some type of functioning adapting to learning very easily, as the picture shows, neural networks have been used in various investigations for data congestion analysis in various types of networks such as ATMs, as well as in the part IT security, obtaining good results. [5]

The discipline in computer science given by the mathematician Alan Turing began the birth of Artificial Intelligence that aims to question whether machines have the ability to think and is currently known as the Turing test. To obtain a good design of the topology of a network with restrictions using the tools of artificial intelligence, it is necessary to build an algorithm that analyses each and every one of the possible solutions. [6]



Figure 2. Intelligent Systems applied to Data [4]

2.3. In the research "Neural networks and traffic prediction":

The use of artificial intelligence tools allows us to predict data traffic, as the picture shows, classify the types of traffic and recognize the existing variables in the analysis of information traffic of a network, These AI methods are additional alternatives that allow us to an easy way to solve this problem, in this case it applies a neural network that based on the weights of neurons and after executing several iterations gives us predictions with data according to expectations. [7]

The different ways that exist to predict traffic consist of extracting the underlying relationships of the previous values and they are also used to extrapolate and predict future behavior. [8]

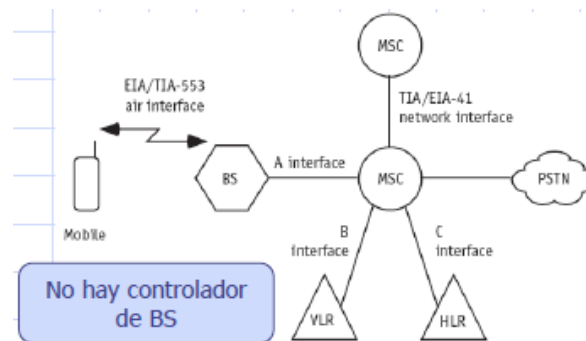


Figure 3. Neural networks and traffic prediction [5]

2.4. In the investigation "Prediction of the demand for fixed telephony through artificial neural networks"

The high increase in users both for fixed telephony and for mobile telephony has made the statistics in bases regarding the demand for the acquisition of these services increases, as the picture shows, the operators that provide this type of service are in need of use tools that provide estimates of next values growth of users and demand for the service. [9]

The adoption of a new technology generally follows a growth pattern of a logistics curve in which low growth with few users is initially identified, followed by high accelerated growth in moderate time intervals. [10]

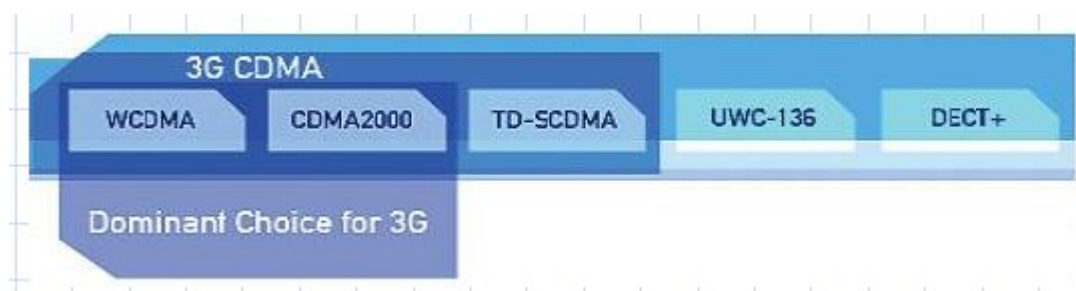


Figure 4. Prediction of the demand for fixed telephony through artificial neural networks [6]

3. WORK METHODOLOGY

Because the proposed objective of the project is the generation of various predictive models of the performance of a base radio third-generation mobile telephony, exactly targeted to traffic of data, based on machine learning algorithms within artificial intelligence, and after an analysis of the techniques existing within this branch and the type of data that the PRS tool (RBS management and monitoring). [11]

- a) For the development and implementation of said models of Predicting the performance of a mobile phone base radio is used the Python Programming Language, due to its variability and the large number of intelligence tools and libraries.
- b) Use of the free platform Google Colaboratory for development of the models, this is a cloud service, which provides us with a Jupyter Notebook that we can access with a web browser without import if we use Windows, Linux or Mac, has as great advantages since being an online tool suitable for this type of jobs offers high-performance technical features such as adequate RAM, possibility of activating a GPU.
- c) A general analysis was performed to determine the properties statistics of the data and especially of the variables to be predicted, when use such information what we determine is the type of data with those that we are going to work on, in addition, it was determined that for the variables to predict there is a very large difference between their values because when compare the maximum, minimum and the mean you can notice that they exist values that can be very small with respect to others of the same variable.
- d) Because the range of the values of the variables to be predicted is extremely broad, it was necessary to include a stage of data normalization in order to avoid a bias in the implementation of the model, after testing and analysing the results, the normalization method known as MinMaxScaler, the same one that generated an optimal normalization of the data with which we should work.[12]
- e) Three models were implemented for each variable and one model final where an estimate of the three variables was attempted at the once with a single model. For this, an order was followed, being the first to always make the selection of the variables to use according to their correlation.
- f) After analysing the results of this last model, they determined that the first variable has no relationship with the other two so when using a single model to estimate all three, the other Variables impair performance in estimating this.

4. DEVELOPMENT

4.1. Obtaining Data

In this part of the work, data was obtained from the behaviour of a third-party mobile phone base radio generation of the Claro operator in Ecuador, it should be noted that.

The data are KPIs of the daily and hourly performance of a radio base of third generation mobile telephony, these data were obtained through the operator's remote monitoring and management tool Sure call PRS. [13]

4.2. Selection Of Variables To Predict

The predictor variables that were selected for the elaboration from this job they were: VS.RAC.DL.EqvUserNum, VS.RAC.UL.EqvUserNum, PS TrafficVolume_GUL, this set of

variables help us to have a general and statistical idea of the behaviour of mobile phone base radios with respect to the data traffic existing in it.

RNC	NodeB Name	Cell Name	Cell ID	CS_TRAFFIC_Volume_GAR_MNO	PS Traffic Volume_GAR (MB)	PS Drop Call_GAR	HSPA User Throughput_GAR	HSPA User Throughput_GAR	VS_RAC_DL_Equipment	VS_RAC_UL_Equipment
RNC004	CAROLINAWS	CAROLINAE	50135	0.0000	0.0000	0.0	0.0	0.0	0.0000	0.0000
RNC004	CAROLINAWS	CAROLINAZ	15139	0.1333	64752.0000	0.0	16450085.0	1818933.0	10222.0000	27145.0000
RNC004	CAROLINAWS	CAROLINAX	15137	0.0333	0.7773	0.0	5652183.0	3394545.0	0.0697	0.1692
RNC004	CAROLINAWS	CAROLINAY	15138	1.3500	17111.0000	0.0	1070702.0	8090640.0	18217.0000	33554.0000
RNC004	CAROLINAWS	CAROLINAG	41571	0.3500	0.1396	0.0	4026316.0	245940.0	0.1390	0.3641
RNC004	SHYRISWS	SHYRISB	52568	0.0000	0.4119	0.0	6322708.0	511043.0	0.0951	0.5048
RNC004	CAROLINAWS	CAROLINAW	15135	0.5000	0.1247	0.0	3036959.0	402147.0	0.3183	11839.0000
RNC004	SHYRISWS	SHYRISF	41513	0.0000	0.0145	0.0	2412.0	400291.0	0.0002	0.0294

Figure 5. Graph of the selection of variables to predict (Google Colaboratory)

4.3. Statistic Analysis

First, a general analysis was carried out to determine the statistical properties of the data and especially of the variables a predict.

Time	RNC	NodeB Name	Cell Name	Cell ID	CS_TRAFFIC_Volume_GAR_MNO	PS Traffic Volume_GAR (MB)	PS Drop Call_GAR	HSPA User Throughput_GAR	HSPA User Throughput_GAR	VS_RAC_DL_Equipment
0 2019-07-08	RNC004	CAROLINAWS	CAROLINAE	50135	0.0000	0.0000	0.0	0.0	0.0	0.0000
1 2019-07-08	RNC004	CAROLINAWS	CAROLINAZ	15139	0.1333	64752.0000	0.0	16450085.0	1818933.0	10222.0000
2 2019-07-08	RNC004	CAROLINAWS	CAROLINAX	15137	0.0333	0.7773	0.0	5652183.0	3394545.0	0.0697
3 2019-07-08	RNC004	CAROLINAWS	CAROLINAY	15138	1.3500	17111.0000	0.0	1070702.0	8090640.0	18217.0000
4 2019-07-08	RNC004	CAROLINAWS	CAROLINAG	41571	0.3500	0.1396	0.0	4026316.0	245940.0	0.1390
5 2019-07-08	RNC004	SHYRISWS	SHYRISB	52568	0.0000	0.4119	0.0	6322708.0	511043.0	0.0951
6 2019-07-08	RNC004	CAROLINAWS	CAROLINAW	15135	0.5000	0.1247	0.0	3036959.0	402147.0	0.3183
7 2019-07-08	RNC004	SHYRISWS	SHYRISF	41513	0.0000	0.0145	0.0	2412.0	400291.0	0.0002
8 2019-07-08	RNC004	SHYRISWS	SHYRISX	17567	0.2167	187625.0000	0.0	16733964.0	706601.0	12401.0000
9 2019-07-08	RNC004	SHYRISWS	SHYRISK	41518	0.0000	0.0406	0.0	3043179.0	978182.0	0.0097

Figure 6. Graph of the data set variables (Google Colaboratory)

Using this information, it was determined that the type of data with those that were worked on, in this case the three variables to be predicted were of the type floating point and 64-bit meaning that they contained a decimal format.

```
[ ] data[['PS Traffic Volume_GUL (MB)', 'VS.RAC.UL.EqvUserNum', 'VS.RAC.DL.EqvUserNum']].info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 21084 entries, 0 to 21083
Data columns (total 3 columns):
PS Traffic Volume_GUL (MB)    21084 non-null float64
VS.RAC.UL.EqvUserNum         21084 non-null float64
VS.RAC.DL.EqvUserNum         21084 non-null float64
dtypes: float64(3)
memory usage: 494.2 KB
```

Figure 7. Graph of the data type of the variables to be predicted (Google Colaboratory)

Furthermore, what was found is that for the variables to be predicted there were a very big difference between their values because when comparing the maximum, minimum and mean it can be noted that there are values that can be very small with respect to others of the same variable. [14]

```
[ ] data[['PS Traffic Volume_GUL (MB)', 'VS.RAC.UL.EqvUserNum', 'VS.RAC.DL.EqvUserNum']].describe()

PS Traffic Volume_GUL (MB) VS.RAC.UL.EqvUserNum VS.RAC.DL.EqvUserNum
count      2.108400e+04      21084.000000      21084.000000
mean       4.749643e+05      86216.887786      33558.809784
std        8.669119e+05      142045.548463      70838.442049
min         0.000000e+00         0.000000         0.000000
25%        5.315750e-01         0.498475         0.055775
50%        8.145250e+04         21640.000000         0.573350
75%        5.724178e+05      100689.500000      32069.250000
max        1.488424e+07      896248.000000      568616.000000
```

Figure 8. Graph of the existing values in the variables (Google Colaboratory)

Moreover, simple graphics prints were also made of the data to visually analyse how the data are distributed and observe if there could be outliers or also known data as outliers.

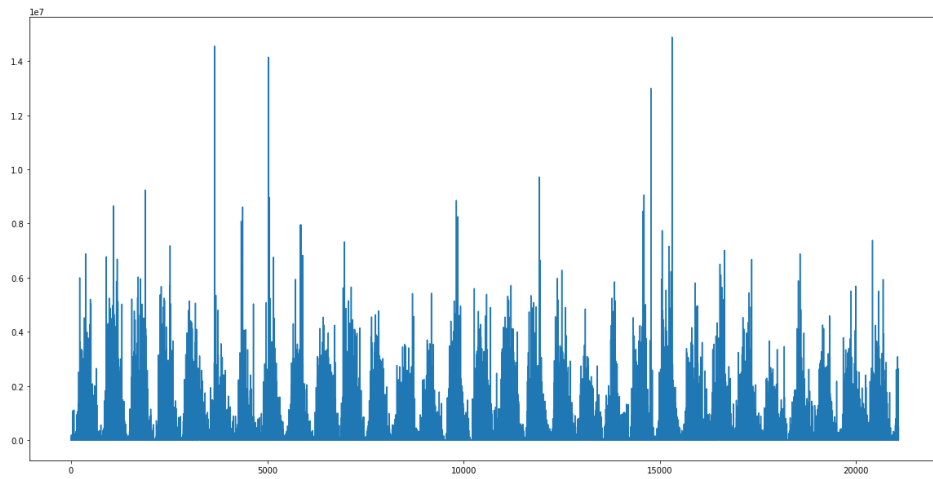


Figure 9. Graph of the data distribution of the variable PS Traffic (Google Colaboratory)

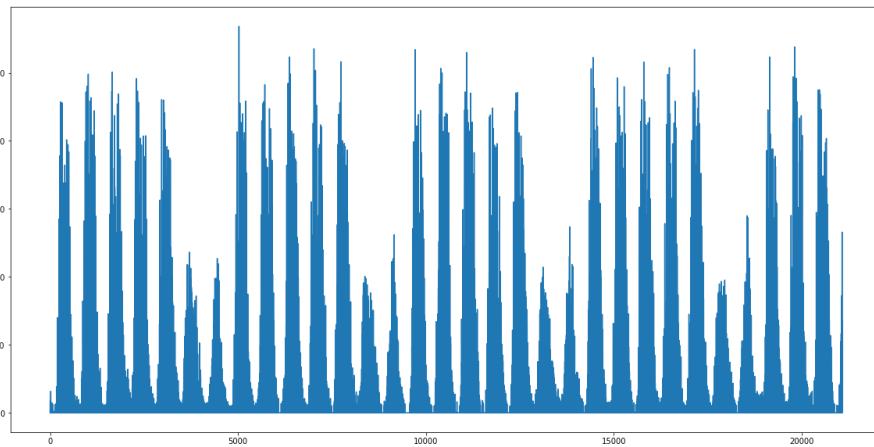


Figure 10. Graph of the data distribution of the variable VS.RAC.DL.EqvUserNum (Google Colaboratory)

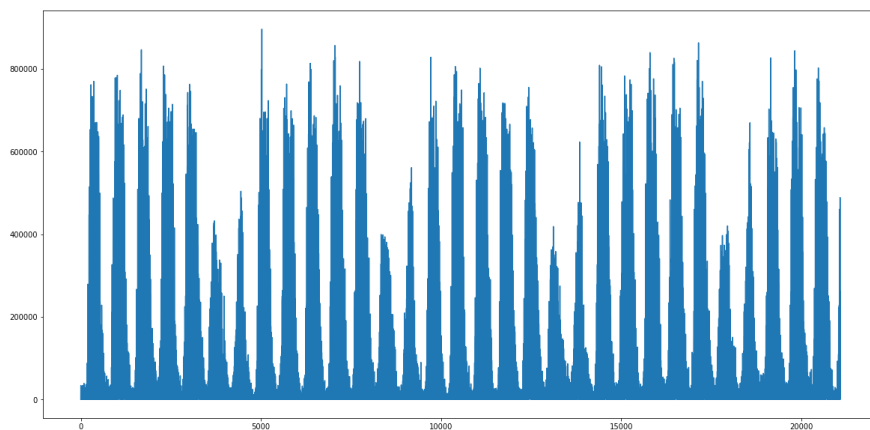


Figure 11. Graph of the data distribution of the variable VS.RAC.UL.EqvUserNum (Google Colaboratory)

4.4. Correlation

Since in the database with which we worked, there was a large number of variables, around a hundred, that could be used to estimate the target variables, to determine the ones with the most information that can contribute, an analysis of the correlation between all the variables with special emphasis on target variables.

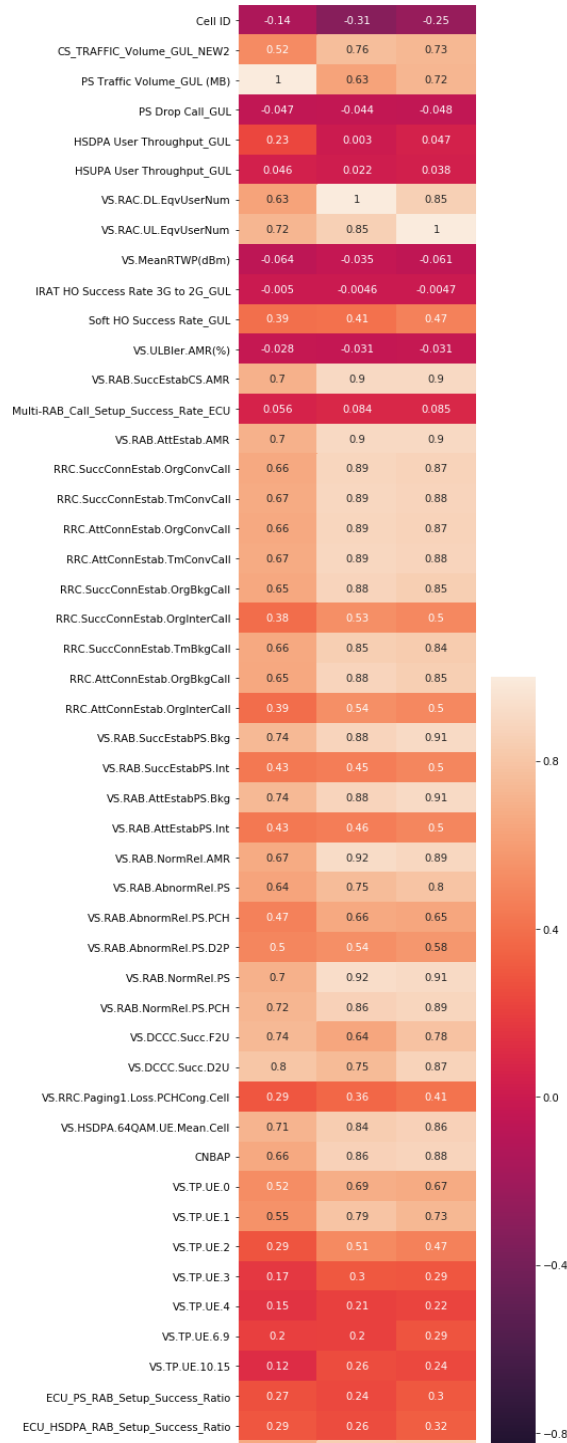
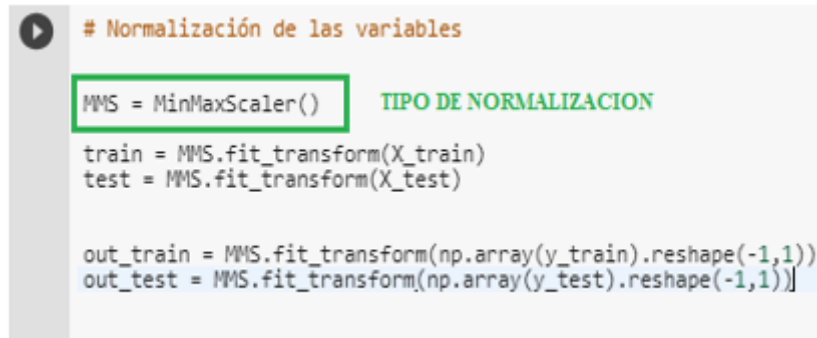


Figure 12. Variable correlation graph (Google Colaboratory)

4.5. Data Normalization

Because the range of the values of the variables to be predicted is extremely broad, it was necessary to include a stage of data normalization in order to avoid a bias in the implementation of the model, after testing and analysing the results, the normalization method known as **MINMAXSCALER**,



```
# Normalización de las variables

MMS = MinMaxScaler()
train = MMS.fit_transform(X_train)
test = MMS.fit_transform(X_test)

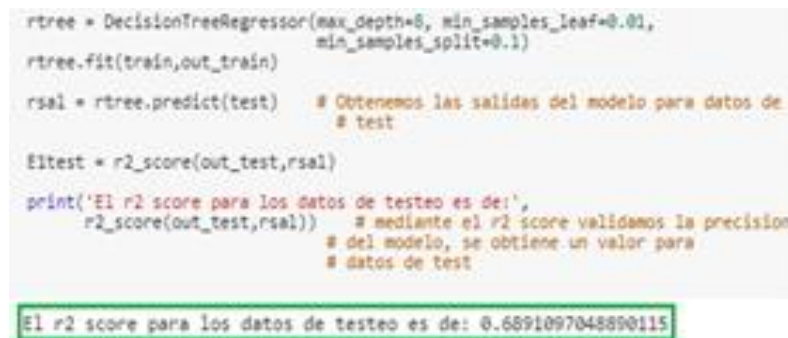
out_train = MMS.fit_transform(np.array(y_train).reshape(-1,1))
out_test = MMS.fit_transform(np.array(y_test).reshape(-1,1))
```

Figure 13. MinMaxScaler normalization graph (Google Colaboratory)

5. MODELS - VARIABLE PS TRAFIC

5.1. Decision Trees

Using decision trees, the hyperparameters by increasing and decreasing depth, parameters with respect to the sheets and others, in order to find the best ones. [13] As we can see in the graph the result that is displayed is the value of r2 score, this value is a metric that tells me the accuracy in prediction calculations of variables that as we can see for this model is a value of 0.6891097048890115



```
rtree = DecisionTreeRegressor(max_depth=8, min_samples_leaf=0.01,
                             min_samples_split=0.1)
rtree.fit(train,out_train)

rsal = rtree.predict(test) # Obtenemos las salidas del modelo para datos de
# test

Eltest = r2_score(out_test,rsal)

print('El r2 score para los datos de testeo es de:',
      r2_score(out_test,rsal)) # mediante el r2 score validamos la precision
# del modelo, se obtiene un valor para
# datos de test

El r2 score para los datos de testeo es de: 0.6891097048890115
```

Figure 14. R2 score graph for the Decision Tree (Google Colaboratory)

5.2. Neural Network

In the same way, for the neural network model, various tests modifying the number of neurons in the hidden layer and the number of hidden layers, as well as their function of activation, its learning rate, its optimization algorithm, number of epochs and overtraining control, looking for optimize the value of your hyperparameters for better performance [15], the optimizer that was used was an LBFGS optimizer, the r2 Score was 0.6979385979700603

```

# creamos el modelo de la NN base
clf = MLPRegressor(hidden_layer_sizes=[(250,200)],activation='identity',max_iter=100000, alpha=0.1, solver='lbfgs',
                  tol=0.00001, learning_rate_init=0.001)

# Realizamos el entrenamiento del modelo
model = clf.fit(train,out_train)

output = model.predict(test)

Eltest = r2_score(out_test,output)
Eltest

r2_score(out_test,output)
0.6970325979700003

```

Figure 15. Graph of r2 score for the Neural Network (Google Colaboratory)

5.3. Random Forest

Finally, with the Random Forest technique what was done was a search using grid search to find your best hyperparameters, prior to this test were performed modifying the parameters manually and finding in which range is where better performance has the model similar to the process used for the other models, the r2 Score was 0.7162649923099472

```

[ ] best_grid = grid_search.best_estimator_
    predictions = best_grid.predict(test)
    r2_score(out_test,predictions)
0.7162649923099472

```

Figure 16. Graph of r2 score for Random Forest (Google Collaboratory)

Once this was determined, a finer adjustment was made. using grid search and using only parameters in ranges where better performance was found with an r2 Score 0.7196744018529605

```

[ ] predictions = regressor.predict(test)
    r2_score(out_test,predictions)
0.7196744018529605

```

Figure 17. R2 score plot for Random Forest with fine adjustment (Google Colaboratory)

A similar procedure was used for the other two variables. During the training and testing of the models, the only thing that changed is that in these no atypical data were found as in the case of the data traffic variable. [16]

5.3.1. Variable VS.RAC.DL.EqvUserNum

a) Decision trees

As we can see in the graph the result that is displayed is the value of r2 score, this value is a metric that tells me the accuracy in prediction calculations of variables that as we can see for this model is a value of 0.8324425441778565

```

rtree = DecisionTreeRegressor(max_depth=5, min_samples_leaf=0.01,
                             min_samples_split=0.01)
rtree.fit(train_DL, out_train_DL)

rsal = rtree.predict(test_DL) # Obtenemos las salidas del modelo para datos de
                             # test

Eltest=r2_score(out_test_DL,rsal)

print('El r2 score para los datos de testeo es de:',
      r2_score(out_test_DL,rsal)) # mediante el r2 score validamos la precision
                                # del modelo, se obtiene un valor para
                                # datos de test

```

El r2 score para los datos de testeo es de: 0.8324425441778565

Figure 18. R2 score graph for the Decision Tree (Google Collaboratory)

b) Neural Network

R2 Score of 0.8883605358098641

```

[] # creamos el modelo de la RNA base
clf = MLPRegressor(hidden_layer_sizes=(100,75),activation='identity',max_iter=10000, alpha=0.1, solver='lbfgs',
                  tol=0.0001, learning_rate_init=0.001)

# Realizamos el entrenamiento del modelo
model = clf.fit(train_DL,out_train_DL)

output = model.predict(test_DL)

Eltest = r2_score(out_test_DL,output)

Eltest

```

/usr/local/lib/python3.6/dist-packages/sklearn/neural_network/multilayer_perceptron.py:112: DataConversionWarning: A y = column_or_1d(y, warn=True)
0.8883605358098641

Figure 19. R2 score graph for the Neural Network (Google Colaboratory)

c) Random Forest

R2 score was 0.8915394414631109

```

[] best_grid = grid_search.best_estimator_
   predictions = best_grid.predict(test_DL)

   r2_score(out_test_DL,predictions)

```

0.8915394414631109

Figure 20. R2 score plot for Random Forest (Google Colaboratory)

Once this was determined, a finer adjustment was made. using grid search and using only parameters with ranges regulated obtaining an r2 Score of 0.8898856746512142

```

[] predictions = regressor.predict(test_DL)

   r2_score(out_test_DL,predictions)

```

0.8898856746512142

Figure 21. Graph of r2 score for Random Forest with fine adjustment (Google Colaboratory)

5.3.2. Variable VS. RAC.UL.EqvUserNum'

a) Decision trees

As we can see in the graph the result that is displayed is the value of r2 score, this value is a metric that tells me the accuracy in prediction calculations of variables that as we can see for this model is a value of 0.9710786186949242. [20]

```
[ ] rtree = DecisionTreeRegressor(max_depth=9, min_samples_leaf=0.01,
    rtree.fit(train_UL, out_train_UL)

    rreal = rtree.predict(test_UL) # Obtenemos las salidas del modelo para datos de
    # test

    Eltest=r2_score(out_test_UL,rreal)

    print('El r2 score para los datos de testeo es de:',
    r2_score(out_test_UL,rreal)) # mediante el r2 score validamos la precision
    # del modelo, se obtiene un valor para
    # datos de test

    El r2 score para los datos de testeo es de: 0.9710786186949242
```

Figure 22. R2 score graph for the Decision Tree (Google Colaboratory)

b) Neural Network

Obtaining an R2 Score of 0.999414359403538

```
[ ] # creamos el modelo de la NN base
    clf = MLPRegressor(hidden_layer_sizes=(100,75),activation='identity',max_iter=10000, alpha=0.1, solver='lbfgs',
    tol=0.0001, learning_rate_init=0.001)

    # Realizamos el entrenamiento del modelo
    model = clf.fit(train_UL, out_train_UL)

    output = model.predict(test_UL)

    Eltest = r2_score(out_test_UL,output)

    Eltest

    /usr/local/lib/python3.8/dist-packages/tensorflow/python/ops/multi_layer_perceptron.py:1111: DataConversionWarning: A column-
    y = column_or_1d(y, warn=True)
    0.999414359403538
```

Figure 23. R2 score graph for the Neural Network (Google Colaboratory)

c) Random Forest

R2 score was 0.9898875480445627

```
[ ] best_grid = grid_search.best_estimator_
    predictions = best_grid.predict(test_UL)

    r2_score(out_test_UL,predictions)

    0.9898875480445627
```

Figure 24. R2 score plot for Random Forest (Google Colaboratory)

Once this was determined, a finer adjustment was made. using grid search and using only parameters in ranges where a better performance was found, and an r2 Score of 0.9917941067924856

```
[ ] predictions = regressor.predict(test_UL)
r2_score(out_test_UL,predictions)
0.9917941067924856
```

Figure 25. R2 score plot for Random Forest with fine adjustment (Google Colaboratory)

5.4. Model to estimate three variables

Once the experiments have been carried out with each variable separately, it was decided to carry out an additional experiment using random forest, since it was the one that best estimated all the variables, to estimate all variables at once and not having an independent model for each variable. In this case, for the selection of variables was chosen for the variables that have a correlation greater than 0.8 with respect to the variables to be predicted. [17]

```
[ ] regressor = RandomForestRegressor()

regressor = RandomForestRegressor(bootstrap=True,
max_depth=100,
max_features=3,
min_samples_leaf=4,
min_samples_split=7,
n_estimators=100)

regressor.fit(train_all, out_train_all)

RandomForestRegressor(bootstrap=True, criterion='mse', max_depth=100,
max_features=3, max_leaf_nodes=None,
min_impurity_decrease=0.0, min_impurity_split=None,
min_samples_leaf=4, min_samples_split=7,
min_weight_fraction_leaf=0.0, n_estimators=100,
n_jobs=None, oob_score=False, random_state=None,
verbose=0, warm_start=False)

[ ]

[ ] predictions = regressor.predict(test_all)

[ ] print('Para datos mb', r2_score(out_test_all[:,0],predictions[:,0]),
'Para datos DL', r2_score(out_test_all[:,1],predictions[:,1]),
'Para datos UL', r2_score(out_test_all[:,2],predictions[:,2]))

Para datos mb 0.541594732750471 Para datos DL 0.8882648157652887 Para datos UL 0.8931188204989436
```

Figure 26. R2 score graph for the three-variable model (Google Colaboratory)

6. RESULTS

6.1. Result of the Model and Testing Variable PS TRAFIC

Regarding the testing process, to validate the models, we used the test data set and the target variable was estimated. Then the exit of the model was compared with the real data using the metric of the r2 score which is widely used in this type of model where regressions are performed, this metric allows us to identify how efficient is the estimate with respect to all the test data.

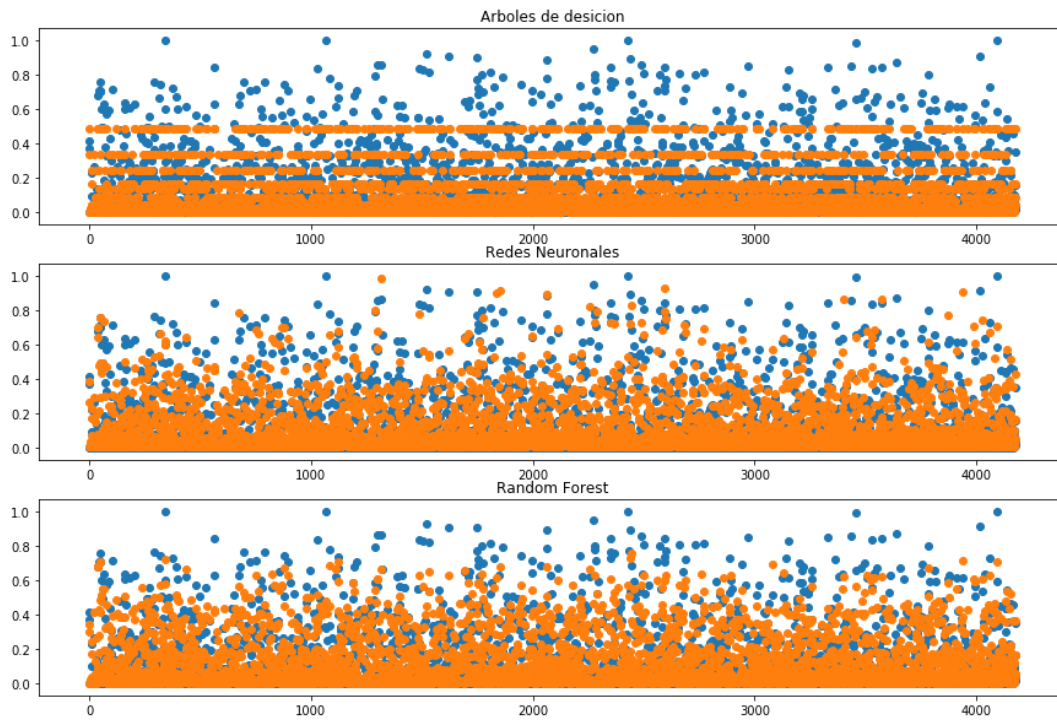


Figure 27. Plot of original and predicted data in each model (Google Colaboratory)

6.2. Outcome of the model and testing variable Vs. RAC.DL.EqvUserNum

Graphs were made of the results obtained and the real data for each model, for said variable, here it could be noted that the decision trees do not have the ability to fully estimate the objective variables, which does not happen with the other models.

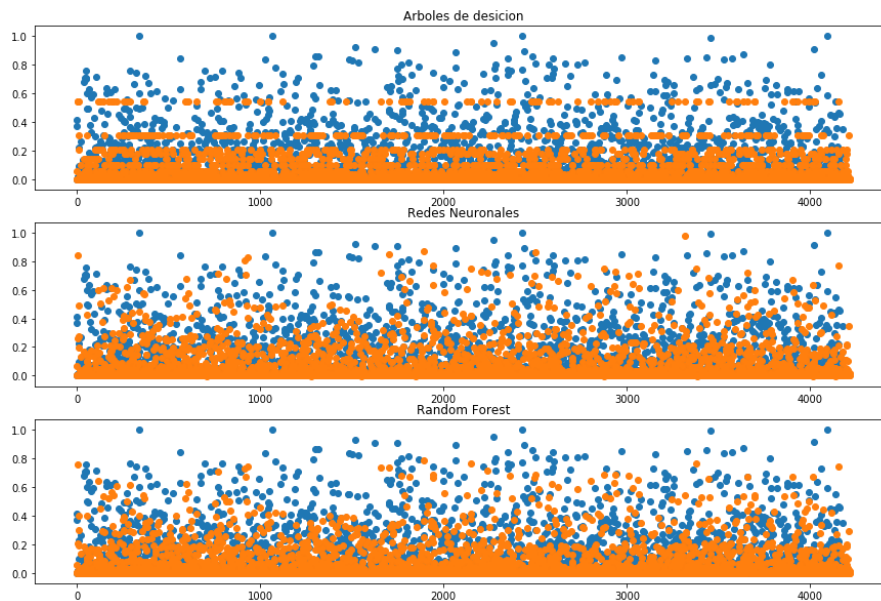


Figure 28. Plot of original and predicted data in each model (Google Colaboratory)

6.3. Outcome of the model and testing variable Vs. RAC.UL.EqvUserNum

Graphs were made of the results obtained and the real data for each model, for said variable, here it could be noted that the decision trees do not have the ability to fully estimate the objective variables, which does not happen with the other models. [19]

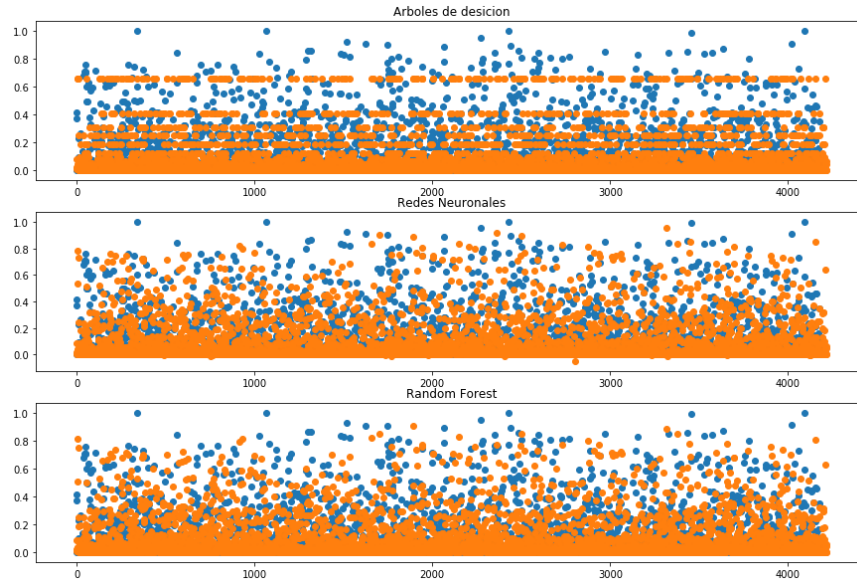


Figure 29. Plot of original and predicted data in each model (Google Colaboratory)

6.4. Discussion and Analysis of Results

As a final part of the development and implementation of the models applied Artificial Intelligence, the numerical analysis of values given in each of the tests, this in order to have a overview and summary of the values obtained, for each of the proposed models and depending on each of the variables of analysis, as detailed below: [18]

Table 1. Analysis of results for all models (Google Colaboratory)

VARIABLE	MODELO	NORMALIZACION	CORRELACIÓN	R2 SCORE
PS TRAFIC MB	Arboles de Decisión	MinMaxScaler	≥ 0.72	0.6891097048890115
PS TRAFIC MB	Redes Neuronales	MinMaxScaler	≥ 0.72	0.6979385979700603
PS TRAFIC MB	Random Forest	MinMaxScaler	≥ 0.72	0.7162649923099472
	Ajuste fino	MinMaxScaler	≥ 0.72	0.7196744018529605
VS.RAC.DL.EqvUserNum	Arboles de Decisión	MinMaxScaler	≥ 0.8	0.8324425441778565
VS.RAC.DL.EqvUserNum	Redes Neuronales	MinMaxScaler	≥ 0.8	0.8883605358098641
VS.RAC.DL.EqvUserNum	Random Forest	MinMaxScaler	≥ 0.8	0.8915394414631109
	Ajuste fino	MinMaxScaler	≥ 0.8	0.8898856746512142
VS.RAC.UL.EqvUserNum	Arboles de Decisión	MinMaxScaler	≥ 0.8	0.9710786186949242
VS.RAC.UL.EqvUserNum	Redes Neuronales	MinMaxScaler	≥ 0.8	0.999414359403538

VS.RAC.UL.EqvUserNum	Random Forest	MinMaxScaler	≥ 0.8	0.9898875480445627
	Ajuste fino	MinMaxScaler	≥ 0.8	0.9917941067924856
Modelo para estimar tres variables	Random Forest	MinMaxScaler	≥ 0.8	PS TRAFIC MB= 0.541594732758471 VS.RAC.DL.EqvUserNum= 0.8881648157651807 VS.RAC.UL.EqvUserNum= 0.8931188204989436

7. CONCLUSIONS

The initial analysis allowed defining which variables shared some type of relationship with the variables we wanted to predict, when using the correlation as a measure of similarity could be filtered to the variables of greater importance and thus use these for the models.

The results were analysed where it was found that the model that the best performance was the model based on random forest, with compared to the other models that were tested, the model based in neural networks was the second that worked best and that of trees decision-maker the one that had the worst performance compared to the others.

It was concluded that neural networks and random forests have had the best results because they have a structure with a greater ability to generalize knowledge.

In the model created to represent all three variables at the same time, the performance of the traffic variable was the one that could not be represented correctly, this behaviour was thought to occur because the other two variables between them share a higher relationship than with that of traffic.

It was concluded that the use of this type of technique is relevant as its use could be extended to applications such as sensors virtual variables, where variables that are complex of measure (due to sensor costs or difficult to access) through variables make them easier to measure.

FUTURE LINES OF WORK

As part of the improvement of this work, it is planned in the future to implement the ARIMA method for the prediction of this type of data, taking into account that ARIMA is a tool that allows working with time series, whether stationary or non-stationary, as we could see at the beginning of our work the data set that was used contains countless variables, in which the variable that relates times or schedules of greater data traffic in an RBS could also be inserted.

In the future this work may be very helpful in the optimization and monitoring of communications networks, whether mobile or fixed, determining their behaviour and performance more quickly.

ACKNOWLEDGEMENTS

The authors thank the higher education institutions that contributed to the research with their authors, as place of work such us: Instituto Superior Tecnológico María Natalia Vaca, Ambato, Ecuador, Instituto Superior Tecnológico Pelileo-Baños, Baños, Ecuador

REFERENCES

- [1] C. Salgado, «Tecnura,» 15 Mayo 2016. [En línea]. Available: <http://www.redalyc.org/jatsRepo/2570/257047577010/index.html>.ERICSSON, «QUALCOMM,»
- [2] E UNIO 2019. [En línea]. Available: <http://itu-apt.org/wrc4p/prez/qualcomm.pdf>.
- [3] E. contributors, «Telefonía móvil 1G,» 17 Enero 2018. [En línea]. Available: https://www.ecured.cu/index.php?title=Telefon%C3%ADa_m%C3%B3vil_1G&oldid=3049358
- [4] Behcet Sarikaya 2000, Packet Mode in Wireless Networks: Overview of Transition to Third Generation, IEEE Comm. Mag. 38 (9): 164-172.
- [5] 4GHS, «4GHS,» [En línea]. Available: <https://www.4ghs.com/4ghs-network-information>.
- [6] Hopcroft John E., Ullman Jeffrey D. 1993 "Introduccion to Automata Theory Languages, and Computation".
- [7] GALIPIENSO, A., ISABEL, M., Cazorla Quevedo, M. A., Colomina Pardo, O., ESCOLANO RUIZ, F. R. A. N. C. I. S. C. O., & LOZANO ORTEGA, M. A. (2003). Inteligencia artificial: modelos, técnicas y áreas de aplicación. Editorial Paraninfo.
- [8] M. Turcanik 2009, "Traffic lights control using recurrent neural networks," Science & Military. vol. 2,
- [9] I. Rohde & Schwarz USA, «Rohde & Schwarz USA, Inc.,» [En línea]. Available: <https://www.mobilewirelesstesting.com/challenges-in-testing-multistandard-radio-base-stations/>.
- [10] BELL, D.K., DE TIENNE, D.H. Y JOSHI, S.A. (2003): «Neural networks as statistical tools for business researchers», Organizational Research Methods, Vol. 6, No 2, pp. 236-265
- [11] J. B. V., «DESARROLLO Y SIMULACIÓN DE UNA ESTACIÓN BASE,» 10 noviembre 2002. [En línea]. Available: <https://scielo.conicyt.cl/pdf/rfacing/v10/art02.pdf>.
- [12] J. Sánchez García, «e-Gnosis,» 1 Enero 2005. [En línea]. Available: <http://www.redalyc.org/articulo.oa?id=73000304>.
- [13] L. C. Corbalán, «SECIDI,» 1 Noviembre 2006. [En línea]. Available: http://sedici.unlp.edu.ar/bitstream/handle/10915/4139/Documento_completo_.pdf?sequence=1&isAllowed=y
- [14] M. COCA, «IESE,» 2007. [En línea]. Available: http://www.iese.umss.edu.bo/uploads/docs/revista_1277232118.pdf#page=153.
- [15] N. S. T. ÁLVAREZ, «TECNURA,» 29 Agosto 2011. [En línea]. Available: <http://www.redalyc.org/articulo.oa?id=257020887009>.
- [16] Olabe, X. B. (1998). Redes Neuronales Artificiales y sus aplicaciones. Publicaciones de la Escuela de Ingenieros.
- [17] O. R. Gámez, «Redalyc,» 1 Marzo 2005. [En línea]. Available: <http://www.redalyc.org/articulo.oa?id=181517913002>.
- [18] Pedro Malagón, Patricia Arroba, Samira Briongos, Alex Mauricio Santana, and José M. Moya. 2020. Modeling tree-structured I2C communication to study the behavior of a dielectric coolant in a two-phase immersion cooling system.
- [19] In Proceedings of the 2020 Summer Simulation Conference (SummerSim '20). Society for Computer Simulation International, San Diego, CA, USA, Article 33, 1–12. <https://dl.acm.org/doi/10.5555/3427510.3427544>
- [20] R. Gámez 2005, «Telefonía móvil celular: origen, evolución, perspectivas,» Ciencias Holguín, p. 9.

AUTHORS

Eduardo Luis Calo Villalva - Master's Degree in Artificial Intelligence (Universidad Internacional de la Rioja), Electronics and Communications Engineer (Technical University of Ambato), teacher in the Electricity career at the Instituto Superior Tecnológico María Natalia Vaca.



Gabriel Alejandro Vaca Ortega - Master's Degree in Systems Management (Universidad de las Fuerzas Armadas ESPE), Electronics and Communications Engineer (Technical University of Ambato), coordinator and teacher in the Electricity career at the Instituto Superior Tecnológico María Natalia Vaca.



Cristina Del Rocio Sanchez Lara (Universidad de Fuerzas Armadas ESPEL - Latacunga), Mechatronics Engineer. Project execution and automation engineer. Instructor of Educational Robotics with Mechatronic Legos. Full-time professor at the Instituto Superior Tecnológico María Natalia Vaca in the automotive area – in the city of Ambato.



David Alejandro Jines Espín. - Master in Technologies for the management and teaching practice (Pontificia Universidad Católica del Ecuador Ambato Headquarters), Electronics and Communications Engineer (Technical University of Ambato). SNNA Leveling Teacher at the Technical University of Ambato. Currently Professor of the electronics career at the Instituto Tecnológico Superior María Natalia Vaca.



Washington Giovanni Amancha Punina. - Master's Degree in Artificial Intelligence (Universidad Internacional de la Rioja), Electronics and Communications Engineer (Universidad Técnica de Ambato). IT Analyst at Tata Consultancy Services (TCS). Professor of Instituto Superior Tecnológico SECAP. I currently work as a coordinator and teacher in the Electronica career at the Instituto Superior Tecnológico María Natalia Vaca.



Angel Arturo Flores Lescano – Master in Management information systems (Universidad Autónoma de los Andes – Ambato), Electronic and Communications Engineer (Universidad Técnica de Ambato), Electronics professor at the Instituto Superior Tecnológico María Natalia Vaca.



Alex Mauricio Santana Gallo. - Master's Degree in Electronic Systems Engineering (Universidad Politécnica de Madrid), Visiting Researcher HCT-LAB Universidad Autónoma de Madrid 2016 (Universidad de Fuerzas Armadas ESPE), Electronics and Instrumentation Engineer (Universidad de Fuerzas Armadas ESPE). CEO of the company IA-KUNTUR S.A.S. B.I.C. dedicated to the development of artificial intelligence and data science located in the city of Latacunga- Ecuador. Currently teaching the electronic career at the Instituto Superior Tecnológico María Natalia Vaca.



María Fernanda Oñate Pico. - Master in Higher Education (Universidad de Fuerzas Armadas ESPE), Teacher in technical specialty (Instituto Superior Tecnológico Pelileo-Baños), Mathematics Teacher (La Salle Ambato Unit), Mechatronics Engineering (Universidad de Fuerzas Armadas ESPE), FINANCIAL MANAGER of the company IA-KUNTUR S.A.S. B.I.C. dedicated to the development of artificial intelligence and data science located in the city of Latacunga-Ecuador.

