**Incremental Learning for SEIRD models in the context of the COVID-19 Pandemic**

**Artículo 1**

**Abstract (jose)**

**1 INTRODUCTION (jose)**

**2. RELATED WORKS (Franklin)**

Debería ser sobre aprendizaje incremental y su uso en tareas de predicción de enfermedades

.

**3. INCREMENTAL LEARNING APPROACHES (Franklin)**

**4. OUR APPROACHES**

**4.1 PROCESS OF VARIABLE DEPENDENCE ANALYSIS  FOR THE SEIRD MODEL FOR INCREMENTAL LEARNING (**yullis**)**

**4.2 OUR INCREMENTAL LEARNING APPROACHES (Edgar+Jose)**

Ensemble learning(see figure #), in general, is a model that makes predictions based on a number of different models. By combining individual models, the ensemble model tends to be more flexible (less bias) and less data-sensitive (less variance). So, in this paper used ensemble methods Bagging, the which consisting in the Training a bunch of individual models in a parallel way, each model is trained by a random subset of the data. These subsets are formed by choosing samples randomly (with repetition) from the training set. The prediction results of each model are validated with the Mean Square Error (MSE) and the model with the lowest MSE is selected.

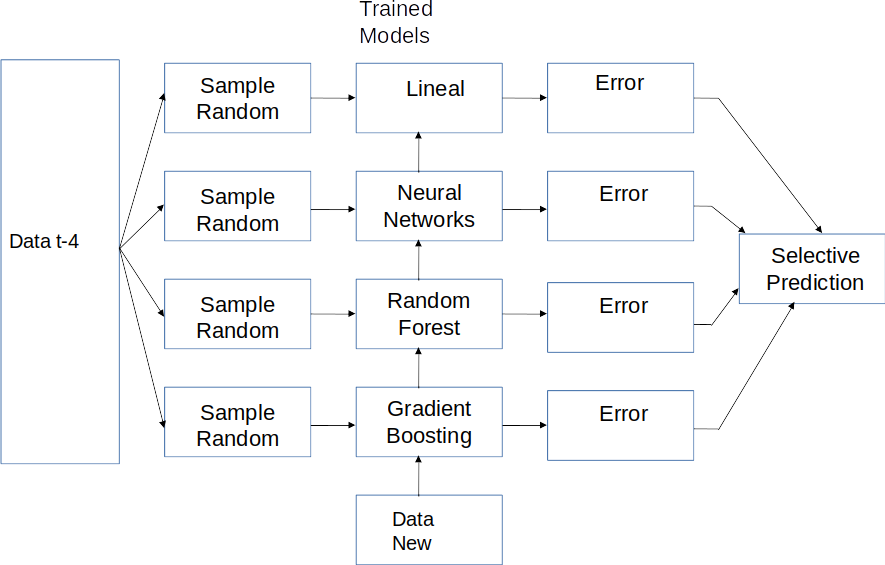


Figure #:Ensemble Learning type Bagging

**5. EXPERIMENTATION**

**5.1 Experimental context (Yullis+Edgar+Douglas)**

(Data sets, casos que se estudiarán, metricas a usar)

A dataset with data for Colombia was used with the following variables per day specifically SEIRD:

* Date: timestamp
* Susceptible: number of susceptible people.
* Exposed: number of people who have exposed.
* Infected: number of people infected.
* Recovered: number of people who have recovered.
* Deaths: number of people who have died.

The data for the variables presented above were obtained from a official source: The National Institute of Health of Colombia (INS), thus guaranteeing the reliability and quality of the data obtained. All the experimentation was done with data that are between March and September of 2020.

For the case study, the target variables were Susceptible, Exposed, Infected, Recovered and Death.

The quality metrics used to measure each model were Mean Square Error (MSE), and coefficient of determination, denoted R2.

**5.2 Experimental Cases**

**5.2.1 VARIABLE DEPENDENCE ANALYSIS (Yullis)**

**5.2.2 Prediction Models (Edgar+Douglas)**

Table 1 shows the performance of random forest incremental predicting the SEIRD variables based on the analysis of the time dependence, where each SEIRD variable has the following dependence according to the results of section 5.2.1:

* Susceptible = exposed(t-5), exposed(t-6), infected(t-5), recovered(t-5), recovered(t-7)
* Exposed = susceptible(t-6), infected(t-7), deaths(t-5), deaths(t-7)
* Infected = exposed(t-6), exposed(t-7), deaths(t-5), deaths(t-6)
* Recovered = susceptible(t-6), susceptible(t-7), exposed(t-5), exposed(t-6), exposed(t-7), infected(t-5), infected(t-7), deaths(t-5)
* Death = susceptible(t-5), exposed(t-5), exposed(t-6), infected(t-5), recovered(t-5), recovered(t-7)

Figures # to #+n shows the performance of random forest incremental, regression linear incremental, neural networks incremental and gradient boosting incremental for predicting the SEIRD variables with the features based on the temporal t-4 analysis of the cross-dependence of the SEIRD variables. Based on these results, each variable has a low error predicting its value, nevertheless, the coefficient of determination is low mainly predicting deaths.

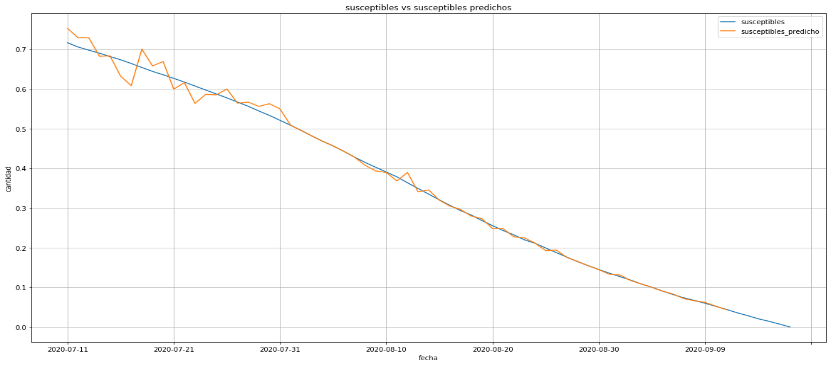


Figure #:Incremental learnings linear t-4 Susceptible

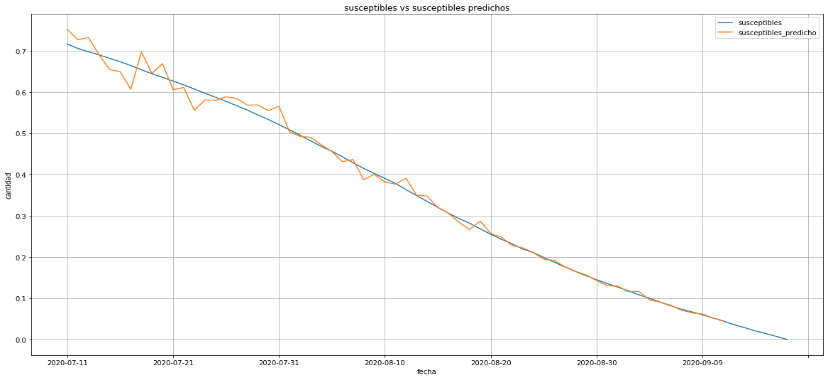


Figure #+1: Incremental learnings Neural Network t-4 Susceptible

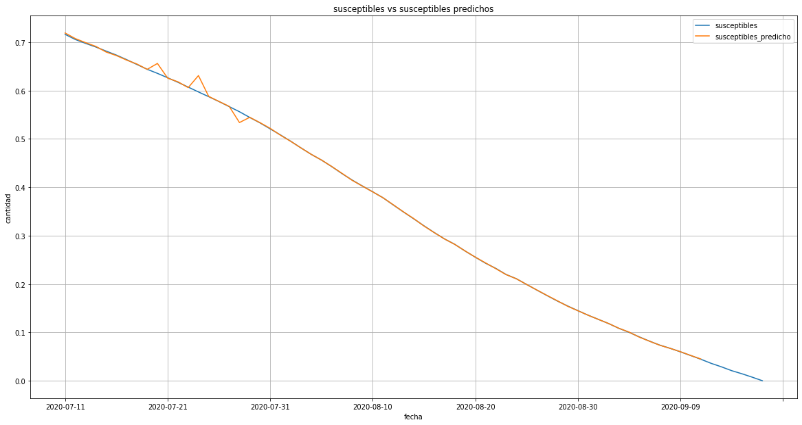


Figure #+2: Incremental learnings Gradient Boosting t-4 Susceptible

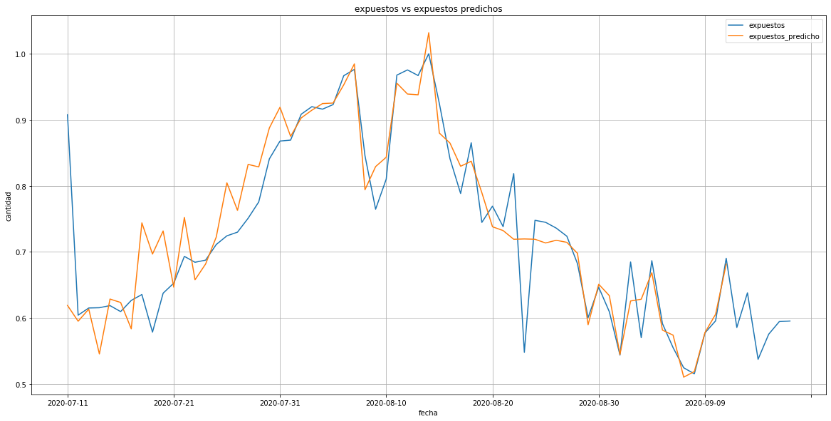


Figure 3+#: Incremental learnings linear t-4 Exposed

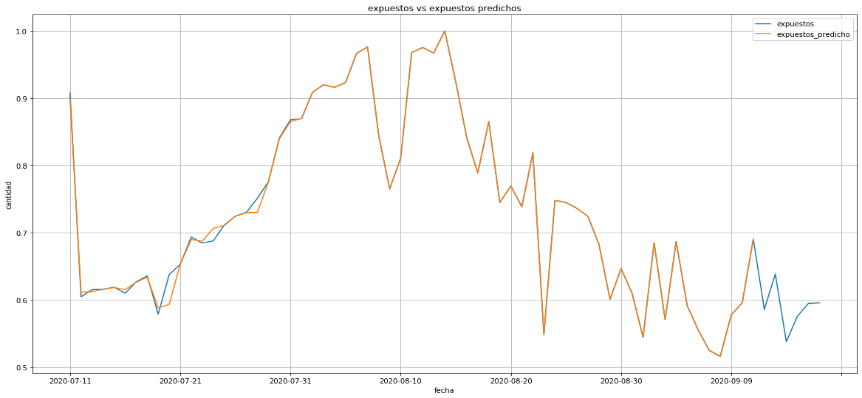


Figure 4+#: Incremental learnings Neural Network t-4 Exposed

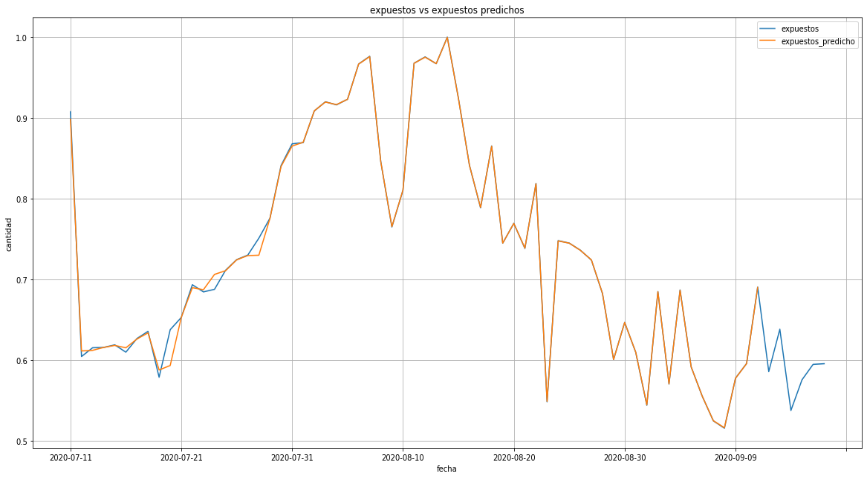


Figure 5+#: Incremental learnings Gradient Boosting t-4 Exposed

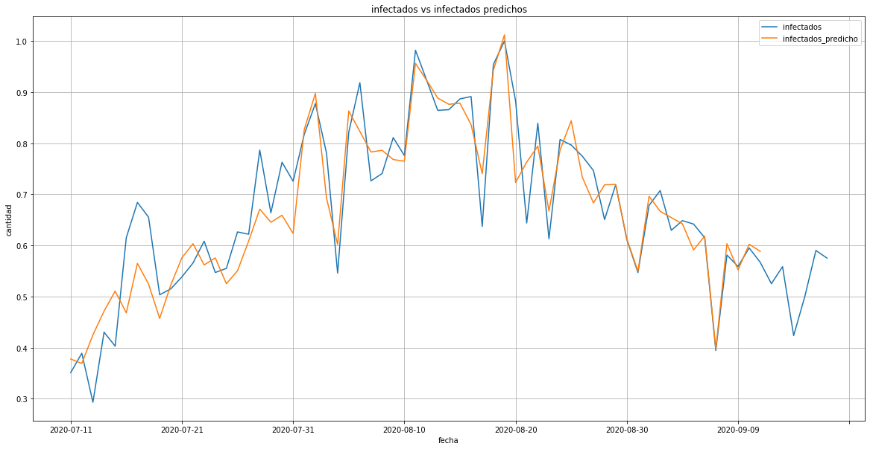


Figure 6+#: Incremental learnings linear t-4 Infectious

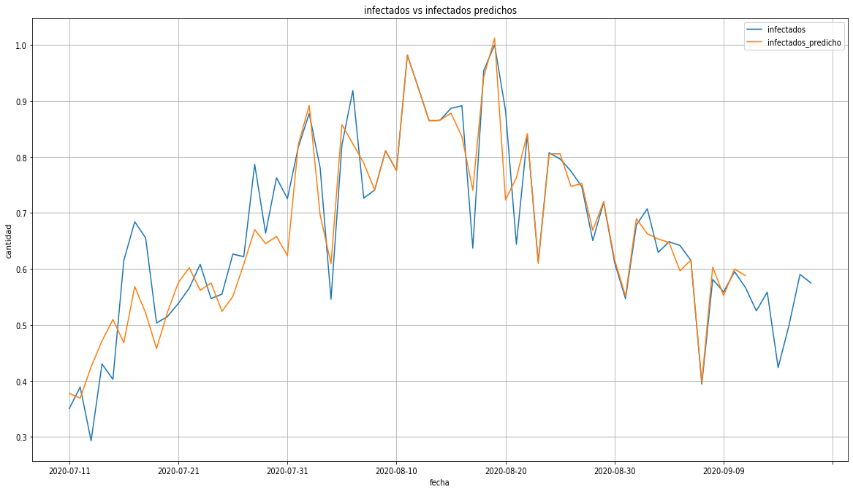


Figure 7+#: Incremental learnings Neural Network t-4 Infectious

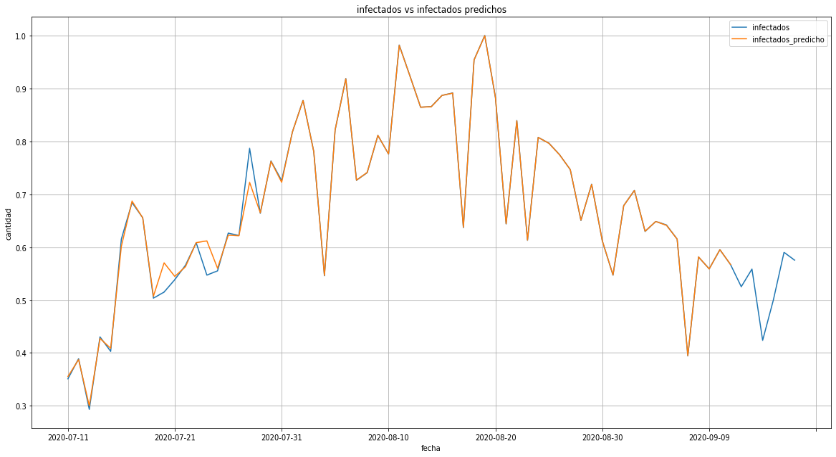


Figure 8+#: Incremental learnings Gradient Boosting t-4 Infectious

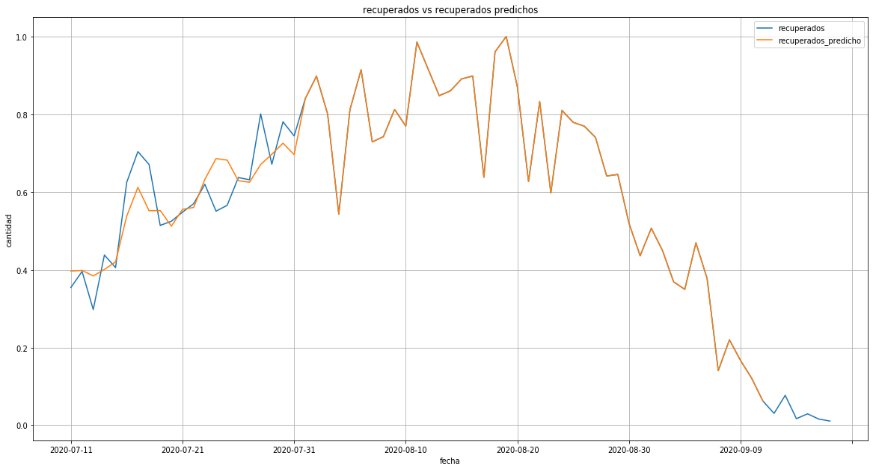


Figure 9+#: Incremental learnings linear t-4 Recovered

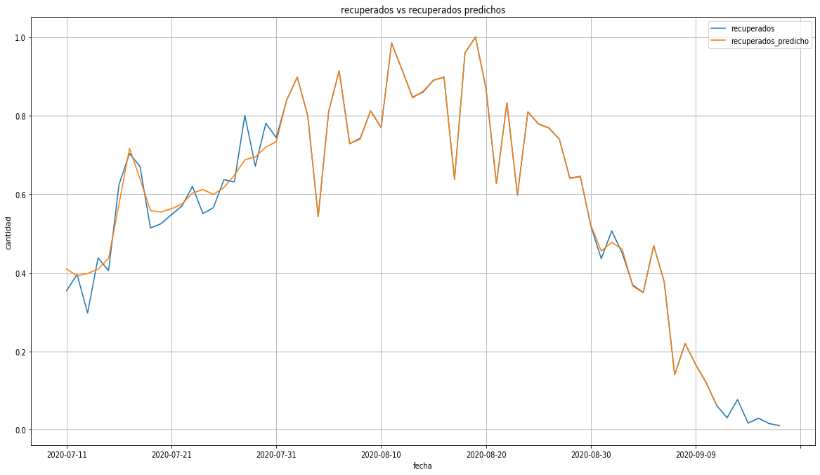


Figure 10+#: Incremental learnings Neural Network t-4 Recovered

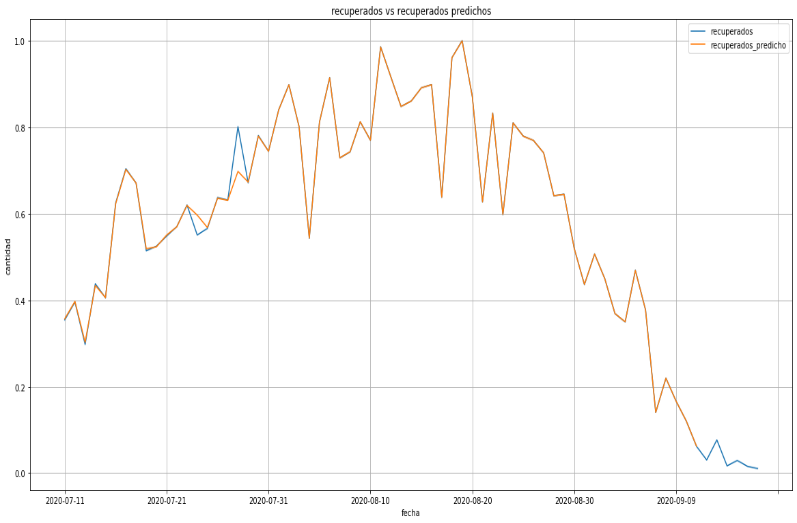


Figure 11+#: Incremental learnings Gradient Boosting t-4 Recovered

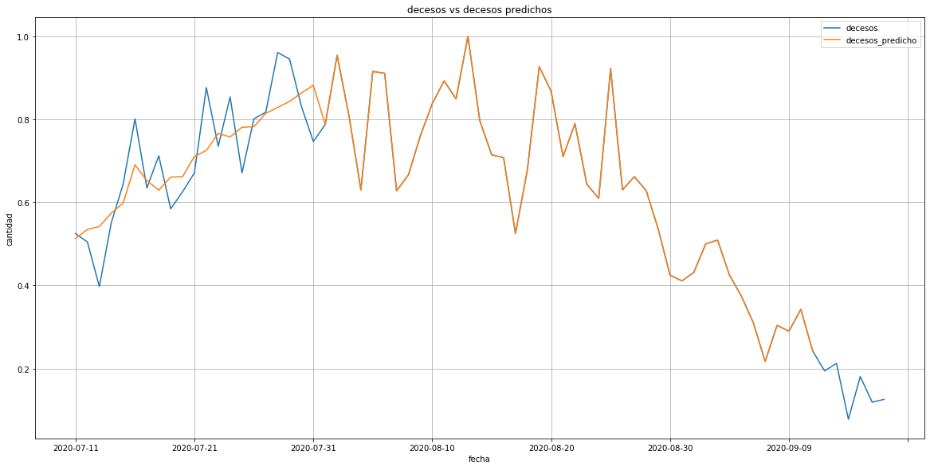


Figure 12+#:Incremental learnings linear t-4 Dead

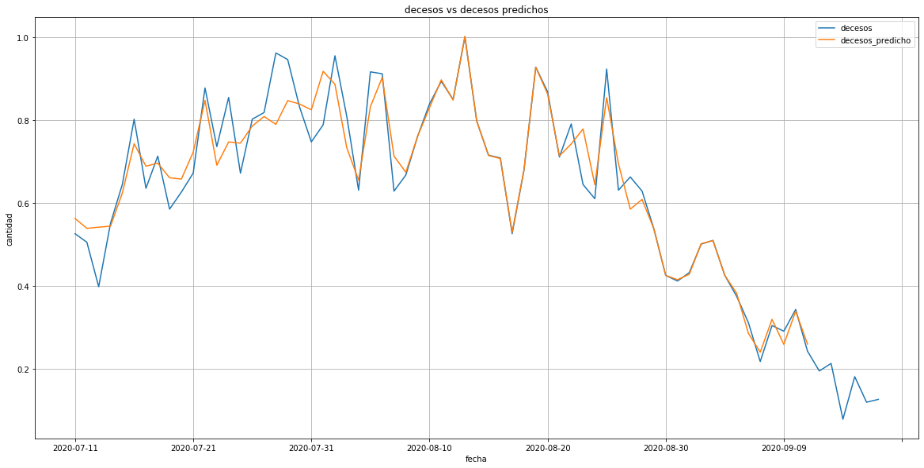


Figure 13+#: Incremental learnings Neural Network t-4 Dead

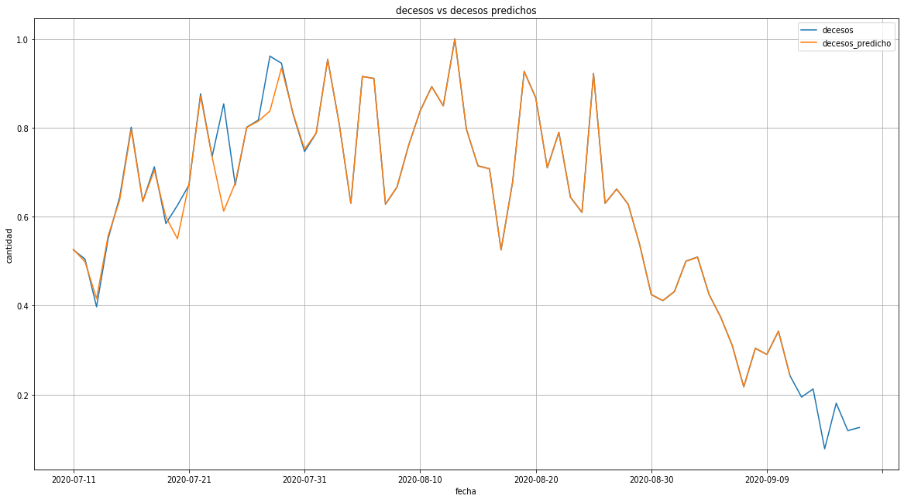


Figure 14+#: Incremental learnings Gradient Boosting t-4 Dead

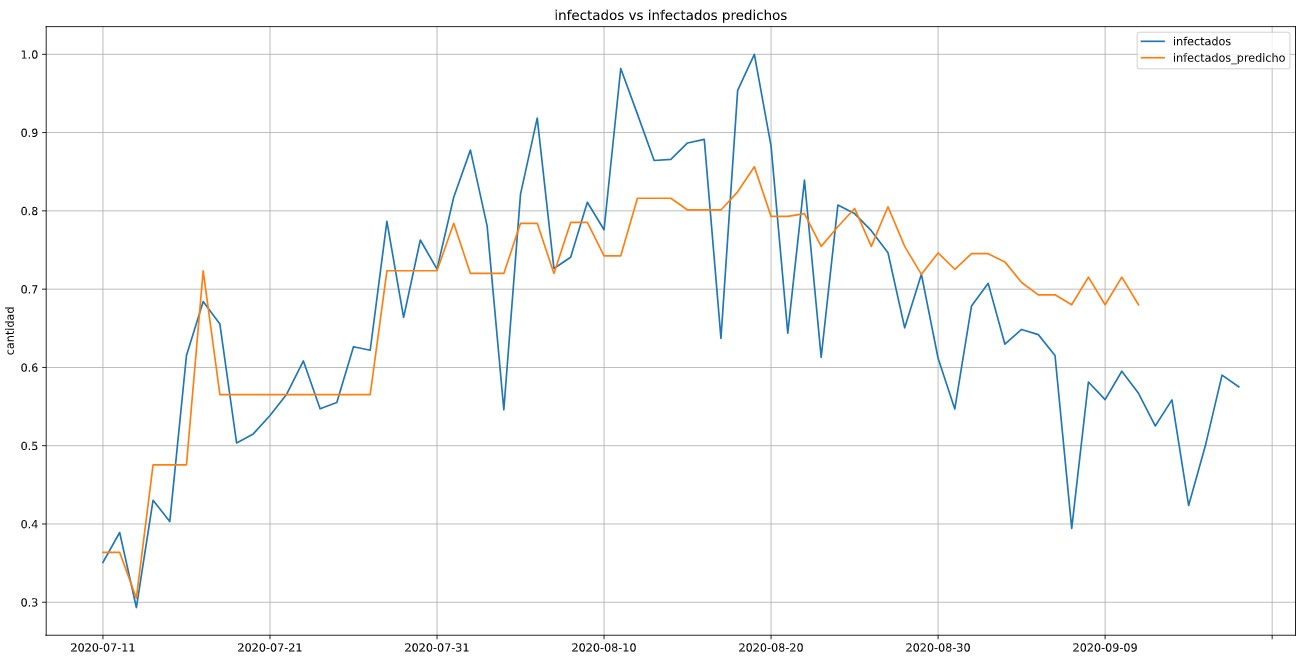


Figure 1. Infected random forest (partial fit)

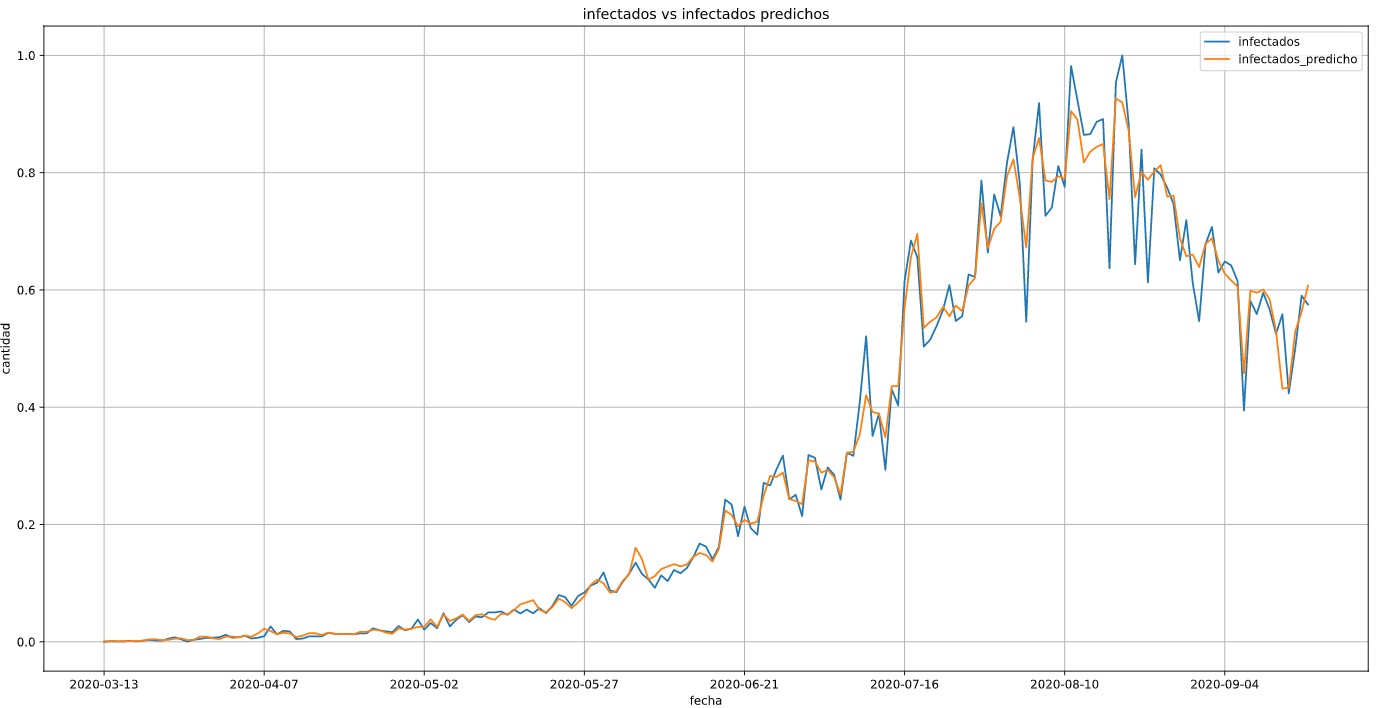


Figure 2 Infected random forest (complete fit)

Table # shows the performance of each model in the training and predicting learning incremental the SEIRD variables based on the analysis of the temporal t-4, with metric used: Mean Square Error (MSE) and coefficient of determination R2.

Table # Quality of the used models to predict learning incremental the SEIRD variables for Colombia

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Target Variable** | **Regressor Model** | **R2** | **Mean Squared Error** | **Incremental Learnings** | **R2** | **Mean Squared Error** |
| **S** | Gradient boosting | 0.990 | 0.0010 | Gradient boosting | 0,999 | 3,1e-05 |
| Random forest | 0,9223 | 0,0078 | Random forest | 0,4220 | 0,0191 |
| Linear | 0,998 | 0.0010 | Linear | 0,9962 | 0.0001 |
| Neural network | 0.9334 | 0.0006 | Neural network | 0,9941 | 0.0001 |
| **E** | Gradient boosting | 0.788 | 0,0023 | Gradient boosting | 31.6654 | 0.9128 |
| Random forest | 0,9422 | 0,0056 | Random forest | 0,2865 | 0,0129 |
| Linear | 0.788 | 0,0023 | Linear | 0,8991 | 0.0018 |
| Neural network | 0.846 | 0.0003 | Neural network | 0,9274 | 0.0013 |
| **I** | Gradient boosting | 0.9718 | 0.0007 | Gradient boosting | 0,0091 | 0.0001 |
| Random forest | 0,9582 | 0,0048 | Random forest | 0,4434 | 0,0103 |
| Linear | 0.824 | 0.0043 | Linear | 0,8400 | 0.0029 |
| Neural network | 0.833 | 0.0053 | Neural network | 0,8626 | 0.0025 |
| **R** | Gradient boosting | 0.979 | 0.0001 | Gradient boosting | 0.9954 | 0.0002 |
| Random forest | 0,9552 | 0,0051 | Random forest | 0,2289 | 0,0411 |
| Linear | 0.8839 | 0.0050 | Linear | 0.9804 | 0.0010 |
| Neural network | 0.9384 | 0.0024 | Neural network | 0.9912 | 0.0004 |
| **D** | Gradient boosting | 0.9185 | 0.0067 | Gradient boosting | 0.9681 | 0.0013 |
| Random forest | 0,9327 | 0,0080 | Random forest | 0,1875 | 0,0355 |
| Linear | 0.829 | 0.0085 | Linear | 0.9591 | 0.0017 |
| Neural network | 0.8625 | 0.006 | Neural network | 0.935 | 0.002 |

**5.3 DISCUSSION OF RESULTS (Todos)**

**6. CONCLUSIONS AND FUTURE WORKS (jose)**

**REFERENCES**