**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.

**5.2.2 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)

Table 1 shows the performance of random forest incremental 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.

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Model | MSE | R2 |
| S | random forest incremental | 0,0191 | 0,4220 |
| E | random forest incremental | 0,0129 | 0,2865 |
| I | random forest incremental | 0,0103 | 0,4434 |
| R | random forest incremental | 0,0411 | 0,2289 |
| D | random forest incremental | 0,0355 | 0,1875 |

Table 1. incremental learning random forest (partial fit)

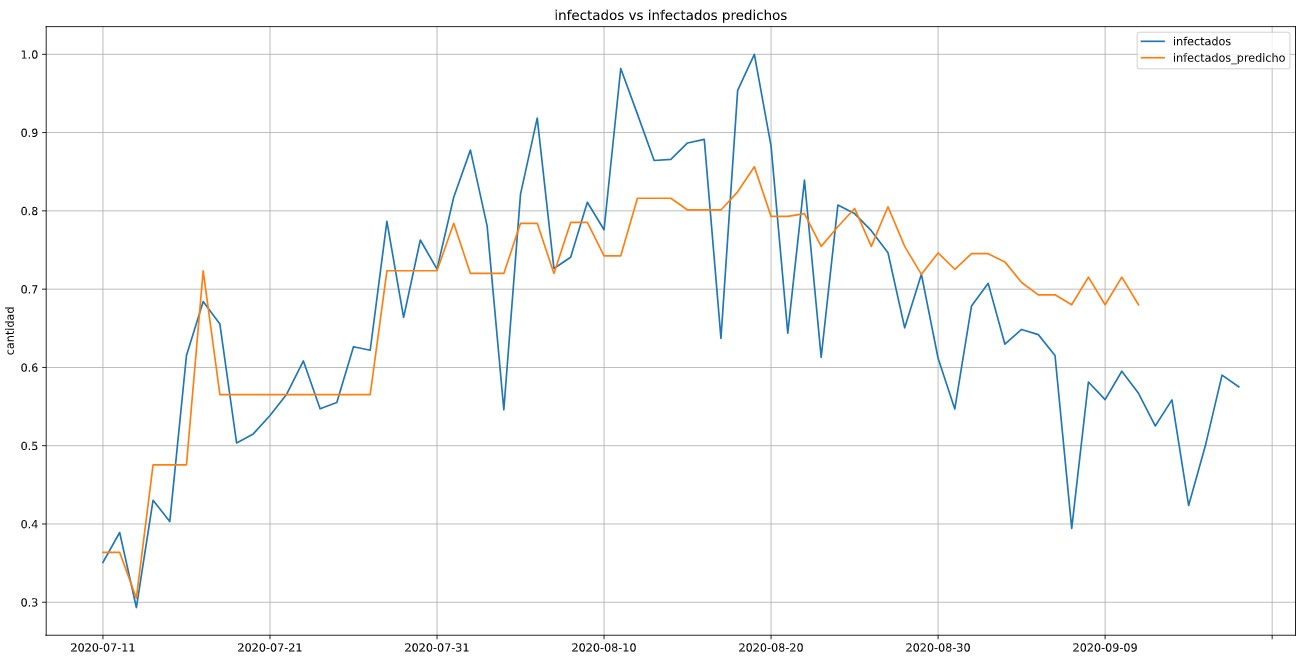


Figure 1. Infected random forest (partial fit)

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Model | MSE | R2 |
| S | random forrest incremental | 0,0078 | 0,9223 |
| E | random forrest incremental | 0,0056 | 0,9422 |
| I | random forrest incremental | 0,0048 | 0,9582 |
| R | random forrest incremental | 0,0051 | 0,9552 |
| D | random forrest incremental | 0,0080 | 0,9327 |

Table 2. incremental learning random forest (fit)

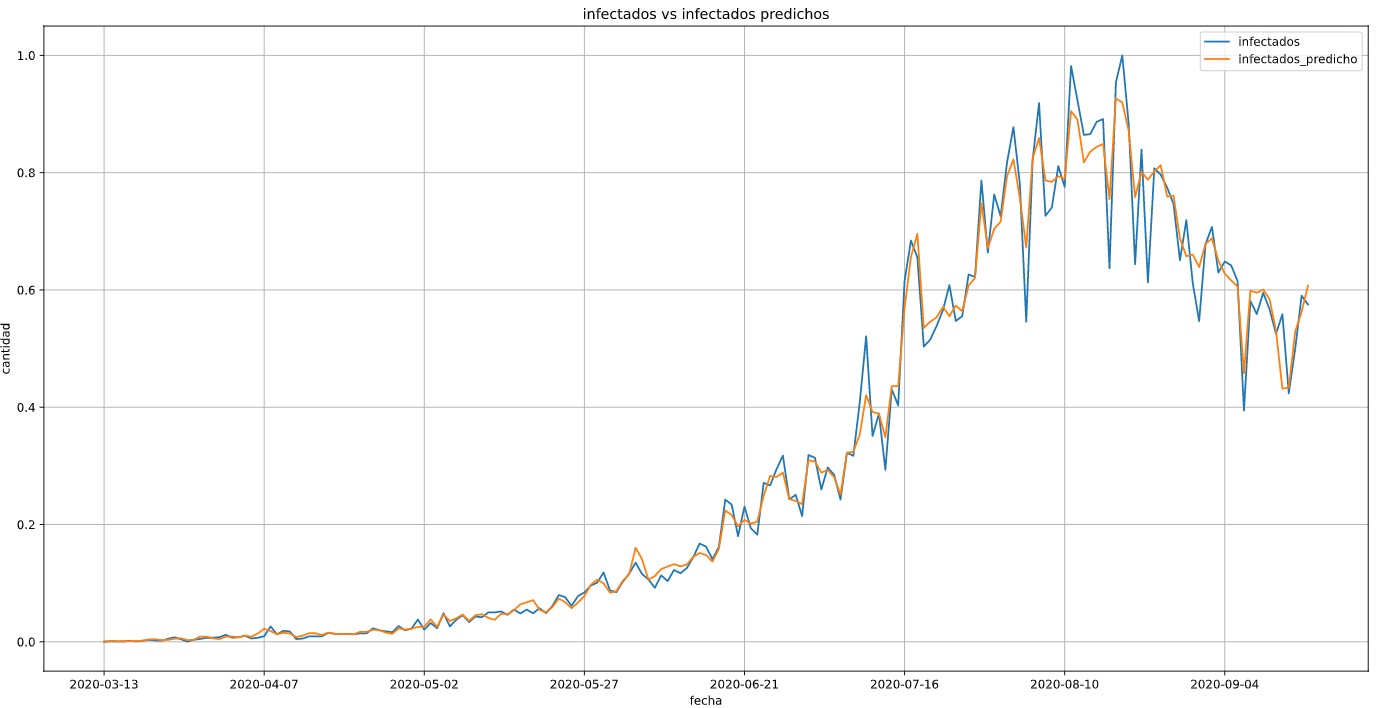


Figure 2 Infected random forest (complete fit)