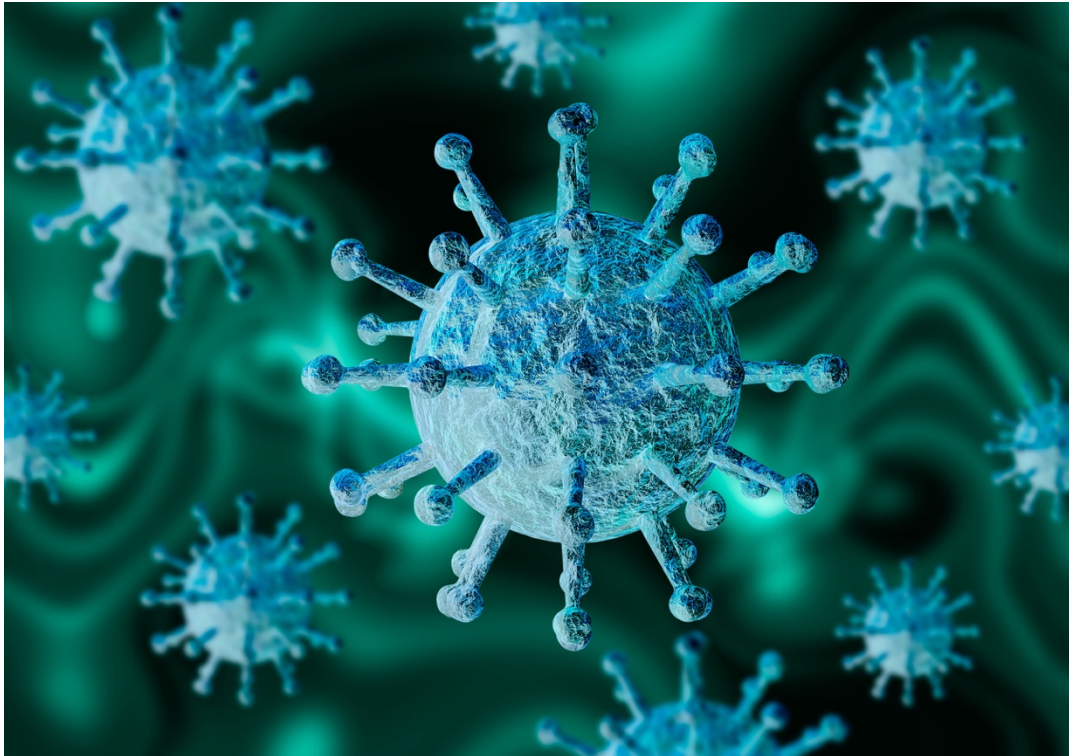# Proyecto 4 – Arboles de decisión COVID-19

DESCRIPCIÓN DEL PROBLEMA

Según informan los científicos, uno de los aspectos claves para controlar la pandemia del COVID-19 es controlar la congestión en los hospitales. Para evitar esta congestión, varias provincias están haciendo pruebas con sistemas que hacen recomendaciones de quedarse en casa o ir al hospital.

Para ello se tienen en cuenta una serie de **factores**:
- Fiebre
- Tos
- Dificultad para respirar
- Fatiga
- Flema
- Dolor muscular
- Dolor de cabeza
- Dolor de garganta
- Escalofríos
- Nauseas
- Congestión nasal
- Diarrea
- Expulsión de sangre a través de la tos
- Conjuntivitis

Nuestra tarea consistirá en:
- Crear el árbol de decisiones 10 veces.
- Determinar el resultado de las 10 ejecuciones.
- Seleccionar los atributos mas representativos.
- Ver el árbol que haya obtenido un resultado mas preciso.
- Testear con los parámetros de la función rpart.
- Dividir aleatoriamente los datos en entrenamiento (80%) y en test (20%).

ANALISIS DE LOS RESULTADOS

- **1º Iteración**

```
[1] "- IT: 1"
Confusion Matrix and Statistics

                 prediction
                  go-to-hospital stay-at-home
  go-to-hospital             314           87
  stay-at-home                52         1146

               Accuracy : 0.9131
                 95% CI : (0.8982, 0.9264)
    No Information Rate : 0.7711
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7618

 Mcnemar's Test P-Value : 0.003929

            Sensitivity : 0.8579
            Specificity : 0.9294
         Pos Pred Value : 0.7830
         Neg Pred Value : 0.9566
             Prevalence : 0.2289
         Detection Rate : 0.1964
   Detection Prevalence : 0.2508
      Balanced Accuracy : 0.8937

       'Positive' Class : go-to-hospital
```

- ## 2º Iteración

```
[1] "- IT: 2"
Confusion Matrix and Statistics

                prediction
                 go-to-hospital stay-at-home
  go-to-hospital            315           86
  stay-at-home              62          1136

               Accuracy : 0.9074
                 95% CI : (0.8922, 0.9212)
    No Information Rate : 0.7642
    P-Value [Acc > NIR] : < 2e-16

                  Kappa : 0.7487

 Mcnemar's Test P-Value : 0.05868

            Sensitivity : 0.8355
            Specificity : 0.9296
         Pos Pred Value : 0.7855
         Neg Pred Value : 0.9482
             Prevalence : 0.2358
         Detection Rate : 0.1970
   Detection Prevalence : 0.2508
      Balanced Accuracy : 0.8826

       'Positive' Class : go-to-hospital
```

- ## 3º Iteración

```
[1] "- IT: 3"
Confusion Matrix and Statistics

                prediction
                 go-to-hospital stay-at-home
  go-to-hospital            325           76
  stay-at-home              56          1142

               Accuracy : 0.9174
                 95% CI : (0.9029, 0.9305)
    No Information Rate : 0.7617
    P-Value [Acc > NIR] : < 2e-16

                  Kappa : 0.7766

 Mcnemar's Test P-Value : 0.09818

            Sensitivity : 0.8530
            Specificity : 0.9376
         Pos Pred Value : 0.8105
         Neg Pred Value : 0.9533
             Prevalence : 0.2383
         Detection Rate : 0.2033
   Detection Prevalence : 0.2508
      Balanced Accuracy : 0.8953

       'Positive' Class : go-to-hospital
```

- **4º Iteración**

```
[1] "- IT: 4"
Confusion Matrix and Statistics

                  prediction
                go-to-hospital stay-at-home
   go-to-hospital            315           86
   stay-at-home              66          1132

              Accuracy : 0.9049
                95% CI : (0.8895, 0.9189)
   No Information Rate : 0.7617
   P-Value [Acc > NIR] : <2e-16

                 Kappa : 0.7428

 Mcnemar's Test P-Value : 0.1233

           Sensitivity : 0.8268
           Specificity : 0.9294
        Pos Pred Value : 0.7855
        Neg Pred Value : 0.9449
            Prevalence : 0.2383
        Detection Rate : 0.1970
  Detection Prevalence : 0.2508
     Balanced Accuracy : 0.8781

      'Positive' Class : go-to-hospital
```

- **5º Iteración**

```
[1] "- IT: 5"
Confusion Matrix and Statistics

                  prediction
                go-to-hospital stay-at-home
   go-to-hospital            332           69
   stay-at-home              58          1140

              Accuracy : 0.9206
                95% CI : (0.9062, 0.9334)
   No Information Rate : 0.7561
   P-Value [Acc > NIR] : <2e-16

                 Kappa : 0.7867

 Mcnemar's Test P-Value : 0.3749

           Sensitivity : 0.8513
           Specificity : 0.9429
        Pos Pred Value : 0.8279
        Neg Pred Value : 0.9516
            Prevalence : 0.2439
        Detection Rate : 0.2076
  Detection Prevalence : 0.2508
     Balanced Accuracy : 0.8971

      'Positive' Class : go-to-hospital
```

- **6º Iteración**

```
[1] "- IT: 6"
Confusion Matrix and Statistics

                  prediction
                   go-to-hospital stay-at-home
  go-to-hospital              332           69
  stay-at-home                 65         1133

              Accuracy : 0.9162
                95% CI : (0.9015, 0.9293)
    No Information Rate : 0.7517
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.7762

 Mcnemar's Test P-Value : 0.7955

            Sensitivity : 0.8363
            Specificity : 0.9426
         Pos Pred Value : 0.8279
         Neg Pred Value : 0.9457
             Prevalence : 0.2483
         Detection Rate : 0.2076
   Detection Prevalence : 0.2508
      Balanced Accuracy : 0.8894

       'Positive' Class : go-to-hospital
```

- **7º Iteración**

```
[1] "- IT: 7"
Confusion Matrix and Statistics

                  prediction
                   go-to-hospital stay-at-home
  go-to-hospital              320           81
  stay-at-home                 46         1152

              Accuracy : 0.9206
                95% CI : (0.9062, 0.9334)
    No Information Rate : 0.7711
    P-Value [Acc > NIR] : < 2.2e-16

                  Kappa : 0.7823

 Mcnemar's Test P-Value : 0.002553

            Sensitivity : 0.8743
            Specificity : 0.9343
         Pos Pred Value : 0.7980
         Neg Pred Value : 0.9616
             Prevalence : 0.2289
         Detection Rate : 0.2001
   Detection Prevalence : 0.2508
      Balanced Accuracy : 0.9043

       'Positive' Class : go-to-hospital
```

- **8º Iteración**

```
[1] "- IT: 8"
Confusion Matrix and Statistics

                prediction
                go-to-hospital stay-at-home
  go-to-hospital            323              78
  stay-at-home              54            1144

              Accuracy : 0.9174
                95% CI : (0.9029, 0.9305)
   No Information Rate : 0.7642
   P-Value [Acc > NIR] : <2e-16

                 Kappa : 0.7759

 Mcnemar's Test P-Value : 0.0453

           Sensitivity : 0.8568
           Specificity : 0.9362
        Pos Pred Value : 0.8055
        Neg Pred Value : 0.9549
            Prevalence : 0.2358
        Detection Rate : 0.2020
  Detection Prevalence : 0.2508
     Balanced Accuracy : 0.8965

      'Positive' Class : go-to-hospital
```

- **9º Iteración (Resultado mas preciso)**

```
[1] "- IT: 9"
Confusion Matrix and Statistics

                prediction
                go-to-hospital stay-at-home
  go-to-hospital            342              59
  stay-at-home              62            1136

              Accuracy : 0.9243
                95% CI : (0.9103, 0.9368)
   No Information Rate : 0.7473
   P-Value [Acc > NIR] : <2e-16

                 Kappa : 0.7991

 Mcnemar's Test P-Value : 0.8557

           Sensitivity : 0.8465
           Specificity : 0.9506
        Pos Pred Value : 0.8529
        Neg Pred Value : 0.9482
            Prevalence : 0.2527
        Detection Rate : 0.2139
  Detection Prevalence : 0.2508
     Balanced Accuracy : 0.8986

      'Positive' Class : go-to-hospital
```

- **10º Iteración**

```
[1] "- IT: 10"
Confusion Matrix and Statistics

                  prediction
                  go-to-hospital stay-at-home
  go-to-hospital               331            70
  stay-at-home                  67          1131

                Accuracy : 0.9143
                  95% CI : (0.8995, 0.9276)
     No Information Rate : 0.7511
     P-Value [Acc > NIR] : <2e-16

                   Kappa : 0.7714

 Mcnemar's Test P-Value : 0.8643

             Sensitivity : 0.8317
             Specificity : 0.9417
          Pos Pred Value : 0.8254
          Neg Pred Value : 0.9441
              Prevalence : 0.2489
          Detection Rate : 0.2070
    Detection Prevalence : 0.2508
       Balanced Accuracy : 0.8867

        'Positive' Class : go-to-hospital
```
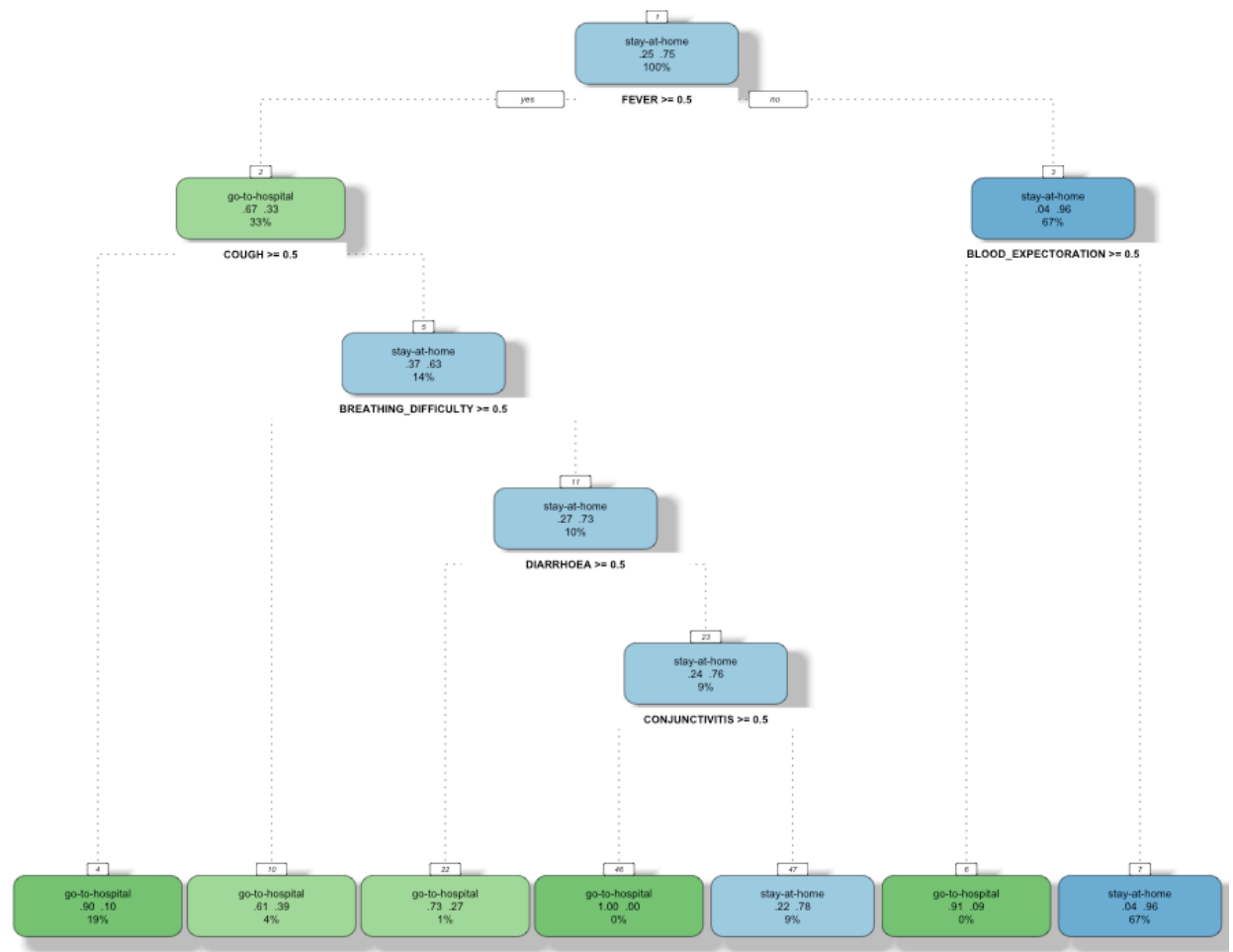
YERAY BELLANCO CASARES

ARBOL DEL RESULTADO MAS PRECISO

RANKING ATRIBUTOS MAS REPRESENTATIVOS

- Fiebre
- Tos
- Expulsión de sangre a través de tos
- Dificultad para respirar
- Diarrea

TEST PARAMETROS FUNCIÓN RPART

TEST 1 (minsplit=1, minbucket=1, maxdepth = 5):

```
Confusion Matrix and Statistics

                prediction
                go-to-hospital stay-at-home
  go-to-hospital            328            73
  stay-at-home              57          1141

              Accuracy : 0.9187
                95% CI : (0.9042, 0.9316)
   No Information Rate : 0.7592
   P-Value [Acc > NIR] : <2e-16

                 Kappa : 0.7807

 Mcnemar's Test P-Value : 0.1883

           Sensitivity : 0.8519
           Specificity : 0.9399
        Pos Pred Value : 0.8180
        Neg Pred Value : 0.9524
            Prevalence : 0.2408
        Detection Rate : 0.2051
  Detection Prevalence : 0.2508
     Balanced Accuracy : 0.8959

      'Positive' Class : go-to-hospital
```

TEST 2 (minsplit=1, minbucket=1, maxdepth = 2):

```
Confusion Matrix and Statistics

                prediction
                go-to-hospital stay-at-home
  go-to-hospital            283           118
  stay-at-home              29          1169

              Accuracy : 0.9081
                95% CI : (0.8928, 0.9218)
   No Information Rate : 0.8049
   P-Value [Acc > NIR] : < 2.2e-16

                 Kappa : 0.7359

 Mcnemar's Test P-Value : 3.925e-13

           Sensitivity : 0.9071
           Specificity : 0.9083
        Pos Pred Value : 0.7057
        Neg Pred Value : 0.9758
            Prevalence : 0.1951
        Detection Rate : 0.1770
  Detection Prevalence : 0.2508
     Balanced Accuracy : 0.9077

      'Positive' Class : go-to-hospital
```

TEST 3 (minsplit=30, minbucket=30, maxdepth = 2):

```
Confusion Matrix and Statistics

               prediction
              go-to-hospital stay-at-home
  go-to-hospital        357           44
  stay-at-home          173         1025

              Accuracy : 0.8643
                95% CI : (0.8465, 0.8807)
   No Information Rate : 0.6685
   P-Value [Acc > NIR] : < 2.2e-16

                 Kappa : 0.6738

 Mcnemar's Test P-Value : < 2.2e-16

           Sensitivity : 0.6736
           Specificity : 0.9588
        Pos Pred Value : 0.8903
        Neg Pred Value : 0.8556
            Prevalence : 0.3315
        Detection Rate : 0.2233
  Detection Prevalence : 0.2508
     Balanced Accuracy : 0.8162

      'Positive' Class : go-to-hospital
```

Como podemos ver en la comparación entre el test 1 y los otros, cuanto mas bajemos el parámetro **maxDepth** menor será la precisión del árbol de decisión debido a que se reduce la profundidad máxima del árbol y por tanto se analizan menos atributos. En el caso del parámetro **minbucket**, podemos ver que en el test3 tiene un valor de 30 mientras que en los otros es de solo 1, cuanto menos sea el numero de minbucket mas divisiones se realizaran en el árbol de decisión y por tanto aumenta la precisión de este. La función del parámetro **minSplit** es la de generar el menor numero de ejemplos que se contemplan en una rama, mientras que minBucket lo hace en las ramas finales.