

# Tarea 1

---

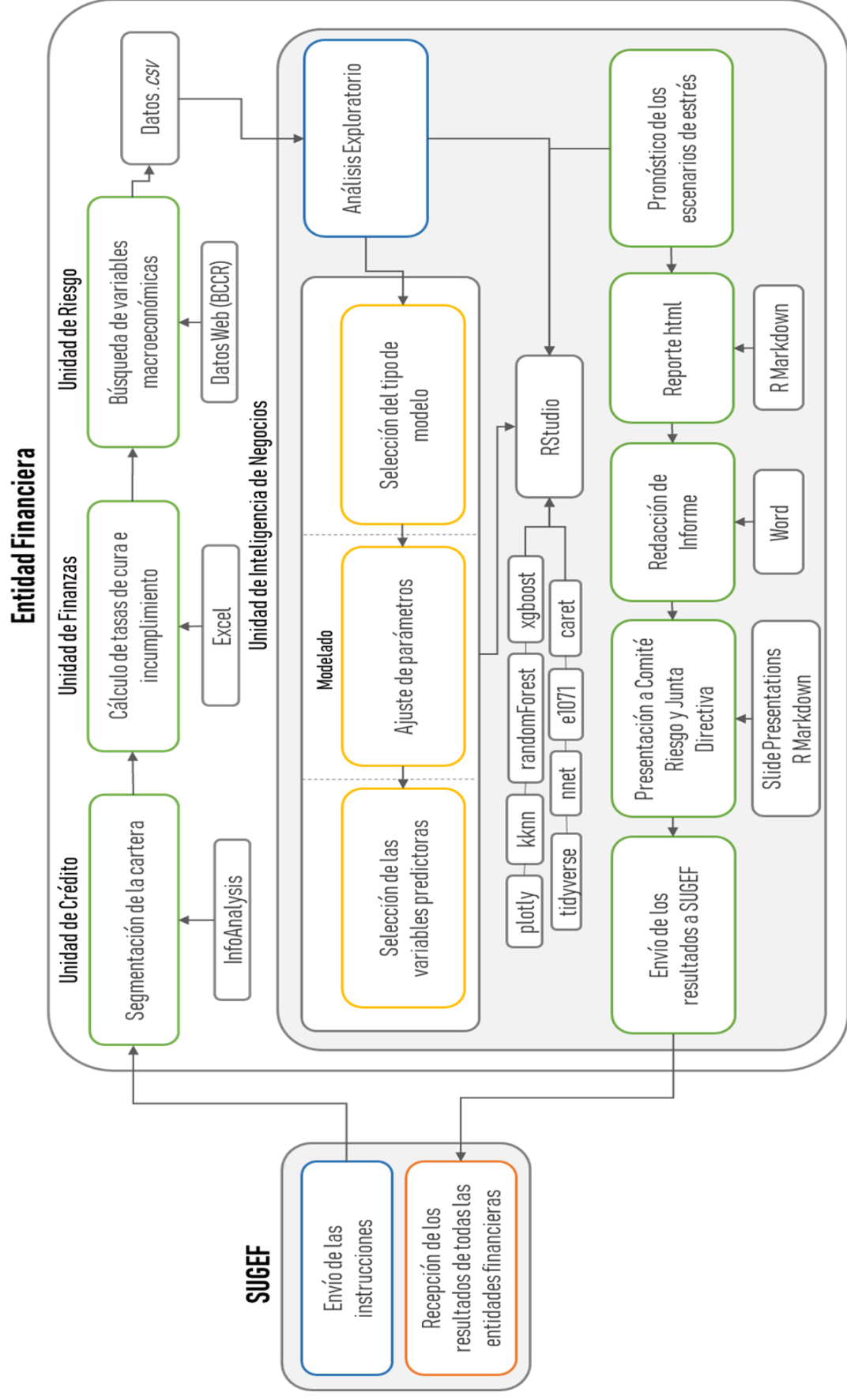
Fiorella Laurito Torres

B23606

SP-1649 Tópicos de Estadística Espacial Aplicada

Setiembre - 2019

# Flujo de trabajo: Modelo Bottom Up Stress Testing



## **Flujo de trabajo: Modelo Bottom Up Stress Testing**

El modelo Bottom Up Stress Testing (BUST) es un proceso anual mandatorio requerido por la Superintendencia General de Entidades Financieras (SUGEF), el cual consiste en una prueba de tensión que tiene el fin de evaluar el rendimiento de las entidades financieras en eventuales escenarios de variables macroeconómicas establecidos por la superintendencia, un escenario base y otro adverso. Mediante la construcción de modelos predictivos para determinar tasas de cura e incumplimiento para los respectivos escenarios.

Para el desarrollo de este proceso es necesario de la participación de distintas unidades de las entidades financieras como: crédito, finanzas, riesgo e inteligencia de negocios; en conjunto trabajan en una serie de pasos para poder concluir con los pronósticos de los escenarios de estrés. El proceso tarda aproximadamente 2 meses y sus respectivas etapas se detallan a continuación.

### **1. Segmentación**

La primera etapa comienza posterior a que SUGEF envíe las instrucciones del ejercicio, se caracteriza por ser una etapa manual ya que todos los clientes de la entidad financiera deben ser clasificados a partir de una segmentación sugerida por la superintendencia, la cual es en función de la actividad económica de los clientes. La segmentación es impartida por la unidad de crédito, a partir del sistema InfoAnalysis donde adquieren información detallada, suministrada por los clientes.

### **2. Recopilación de la información**

Con la clasificación de los clientes establecida, se procede con la recopilación de información requerida para los modelos de pronóstico tanto de las variables predictoras como de las variables respuesta. Esta tarea es asignada a la unidad de finanzas y la unidad de riesgo, esta función la logran mediante macros en Excel y recolección de información de datos presentes en la web, específicamente de información del Banco Central de Costa Rica. Los datos son recopilados en un archivo “.csv” para proceder a analizarlos.

### **3. Modelado**

Esta etapa está a cargo de la unidad de inteligencia de negocios, en la cual impartí mi labor como estadística. El modelado fue realizado para siete segmentos de la cartera (asignados en primera

etapa), respectivamente a la tasa de cura y la tasa de incumplimiento, por lo cual se cuenta con 14 variables respuesta por modelar.

La metodología de análisis consistió en primera instancia un análisis descriptivo de los datos para comprender el panorama de estudio y determinar posibles valores atípicos.

Se procedió con el modelado de los datos, el cual se constituye de tres etapas: la primera consiste en la selección del tipo de modelo predictivo (KNN, redes neurales, arboles de regresión, bosques aleatorios, regresión de soporte vectorial, XGBoost y regresión lineal), en la segunda etapa se realiza un ajuste de parámetros dependiendo del tipo de modelo seleccionado y la tercera etapa se realiza una selección de variables (se probaron todas las posibles combinaciones de las variables). El criterio para seleccionar el mejor modelo en cada una de las etapas, fue la estimación del error cuadrático medio mediante validación cruzada con 10 divisiones de los datos (folds) y 10 repeticiones para analizar la estabilidad del error.

Cabe destacar que el análisis de esta etapa fue realizado a partir de RStudio, con los paquetes: *tidyverse*, *plotly*, *kkn*, *nnet*, *randomForest*, *rpart*, *e1071*, *xgboost* y *caret*. Además, dado que se cuenta con una serie de distintos segmentos, se automatizó mediante funciones de R, específicamente el procedimiento de selección del mejor modelo mediante la estimación del error por validación cruzada. Así de igual manera se efectuaron funciones para graficar y tabular datos de manera automatizada para cada uno de los segmentos.

#### **4. Pronóstico de escenarios de estrés**

Con los modelos predictivos seleccionados, se procedió al pronóstico de los escenarios a partir de las variables macroeconómicas. Estos pronósticos son analizados por todas las unidades involucradas en el proceso para concretar aspectos indispensables para la toma de decisiones.

#### **5. Reportería y documentación**

Para garantizar la reproducibilidad del análisis empleado, se efectúa un reporte *html* en *R-Markdown*, en el cual se detalla todo el código detrás del modelado. Además, se documentó el procedimiento metodológico a la superintendencia, donde se especifica las técnicas y la metodología implementada, para así asegurar la transparencia del mismo. Dicha documentación fue realizada mediante Microsoft Word.

## 6. Aprobación y envío

Para poder proceder con el envío de los resultados a SUGEF, es necesario que el procedimiento sea puesto a revisión por parte del Comité de Riesgo y la Junta Directiva de la entidad financiera, los cuales concretan la aprobación o rechazo del procedimiento. Para esto, las unidades involucradas efectúan una presentación mediante con *Slide Presentations R Markdown*. En caso de refutar el procedimiento, se deben corregir los aspectos solicitados, mientras que, de ser aprobado se sigue con el envío de los resultados a la superintendencia.

## Clasificación del flujo de trabajo

Por último se puso en retrospectiva el flujo de trabajo, por lo cual a continuación se presentan los seis principios, descritos en el blog post: <http://blog.revolutionanalytics.com/2010/10/a-workflow-for-r.html>, donde se especifica si el principio se aplica o no en el flujo de trabajo.

- Transparencia: se realiza una debida documentación donde se especifican todos los pasos y los métodos implementados.
- Capacidad de mantenerlo: cada línea del código está debidamente comentada, además dado que se realiza anualmente se mantiene el formato y nombres de las variables.
- Modularidad: no aplica.
- Portabilidad: no aplica.
- Reproducibilidad: se realizan documentos auto-reproducibles (R Markdown) para garantizar este principio.
- Eficiencia: la gran mayoría del proceso está automatizado mediante funciones de R, esto ya que además de tener que realizar el modelo a varios segmentos, este proceso se repite cada año, lo cual agiliza el procedimiento.