



Proyecto Final


Ernesto Rivera
Big Data

Datos

Sociedades Administradoras de Fondos de Inversión - SAFIs

SUGEVAL posee reports de utilidades y carteras financieras





The way to get started
is to quit talking and
begin doing.

Walt Disney




Yo quiero hacer plata

Un compañero de clase



Agenda



Preprocesamiento
Procesamiento
Resultados

Pre-procesamiento

1

Transformación de tipos de datos desde su lectura, y se omiten columnas.



2

No existen datos nulos o faltantes, solo una SAFI sobrante.



3

De códigos textuales a numéricos.



4

Montos a colones

5

Preparación de la serie de tiempo (t, t+1)

Procesamiento

1

Join
dataframes

2

Vector
Assembler

3

Standard
Scaler

4

Entrenamiento
- Pruebas

5

Modelo de
predicción



Resultados

Análisis



Reflexiones sobre...

Series de tiempo

- Orden importa
- Entrenamiento y pruebas: respetar el orden
- El modelo se debe actualizar cuando se reciban nuevos datos

SUGEVAL

- Datos de buena calidad
- Pero ¿solo en Excel?
 - Openpyxl 📄

Resultado final

Problema de Regresión:

- Árbol de decision
 - $RMSE = 0.006$
- Random forest
 - $RMSE = 0.031$

Resultado final concuerda con los datos reales

*Root mean squared error



cod	p
INSBANCREDITSFI	0.42807087302207947
BNSFI	0.36810361162373223
INTERFINSFI	0.3424972949536026
MUTUALSFI	0.3246776769372905
BCRSFI	0.32467767693729044

Decision Tree



cod	mes	utilidad
INSBANCREDITSFI	2022-03-01	0.42624912
BNSFI	2022-03-01	0.37036204
INTERFINSFI	2022-03-01	0.34626266
MUTUALSFI	2022-03-01	0.32133445
BCRSFI	2022-03-01	0.3198345

Datos Reales

cod	p
INSBANCREDITSFI	0.37643202236030926
BNSFI	0.3430753381280745
INTERFINSFI	0.31443630196909134
BCRSFI	0.29649553968874626
POPULARSFI	0.29291251887027436

Random Forest



Gracias

INS-Inversiones SAFI SA 