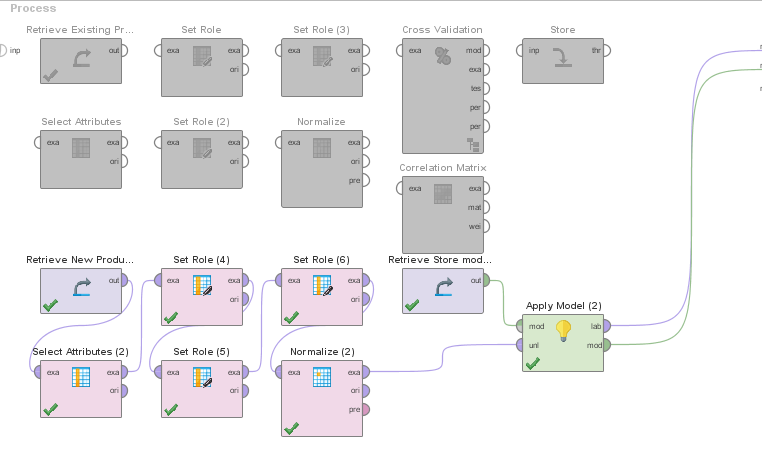
El modelo utilizado fue de predicción por medio de K-NN, ya que este modelo presento los mejores márgenes de RMSE y R-Squared al igual que algunos márgenes de errores. Los otros modelos me daban márgenes menos adecuados al igual que la predicción no es lo que se esperaba. Comparando los datos que ya tenemos de volumen de productos existentes se notan patrones parecidos a los predichos, lo cual da mayor veracidad a los números.

Para correr el modelo primeramente se filtran los diferentes atributos con “Select Attributes” en este caso los atributos seleccionado fueron: PositiveServiceReview, ProductType, ProductNumber, Volume, x5StarReview y x4StarReview. Estos fueron seleccionados por tener mayor correlación entre sí.  
Después se generaron 3 diferentes roles con “set role” El principal genera un Label al Volume, el segundo se le da el rol de ID al ProductType, se hace lo mismo con ProdcutNumber al tercer rol.  
Luego para estandarizar todos los montos, se utiliza “Normalize”.  
Antes de generar una predicción, se debe de seleccionar los datos que sean útiles, ya que si se deja muchos datos la predicción va a ser menos certera de lo esperado, por medio de “Correlation Matrix” se puede medir la correlación entre los datos para usar los mas óptimos.   
Se deja de lado las correlaciones luego de haber tomado una decisión y se introduce el “Cross Validation” el cual dará la predicción y al mismo tiempo medirá los diferentes márgenes de los resultados. Los datos serán entrenados con K-NN. Las pruebas serán con “Apply Model” y “Perfomance Regression”. Cabe destacar que K-NN tenía los K en 3, ese fue el número que dio promedios de error más bajos.   
Por último se utiliza “Store” para almacenar el modelo, dando fin al primer proceso.

Luego de haber reproducido los datos para que se almacenaran se deshabilita todo el primer proceso.  
Se despliegan los datos del cual se desea predecir sus rentabilidad, a estos datos se les debe de depurar de la misma manera que el primer proceso, el cual se quitó de funcionamiento. Con excepción del “Cross Validation” El cual debe ser cambiado por solo “Apply Model”. Este debe llevar una conexión desde el “Normalize” (este trae todo el proceso de depuración a los datos) al “uni” y desde la recuperación de los datos almacenados a “mod”. Luego de reproducir el modelo para ver los resultados se podrá apreciar las predicciones.



**root\_mean\_squared\_error  
root\_mean\_squared\_error: 598.001 +/- 843.417 (mikro: 1033.904 +/- 0.000)**

**root\_relative\_squared\_error  
root\_relative\_squared\_error: 0.425 +/- 0.247 (mikro: 0.710)**

absolute\_error  
absolute\_error: 275.650 +/- 337.633 (mikro: 275.650 +/- 996.481)

relative\_error  
relative\_error: 86.50% +/- 53.84% (mikro: 86.99% +/- 170.49%)

correlation  
correlation: 0.947 +/- 0.066 (mikro: 0.834)

squared\_correlation  
squared\_correlation: 0.900 +/- 0.118 (mikro: 0.696)

