**Cuales aspectos del proceso que siguió funcionaron bien y cuáles no?**

El proceso utilizado, durante su creación tuvo varias direcciones erróneas tomadas al momento, con el afán de presentar los datos de manera más óptima, durante la primera tarea que se trataba de buscar diferentes tendencias entre los usuarios, varias veces el rol elegido como Label no fue adecuado y presentaban datos inservibles para contestar las preguntas a responder, siempre se debe mantener el enfoque hacia a que se desea llegar. Discretizando ciertos atributos relevantes se podía visualizar otras tendencias de compras al igual que generaba cambios en la predicción por medio del Decision Tree. Los operadores debían ser correctamente seleccionados para poder crear el cambio necesario en el modelo, el saber la conexión entre ellos es vital para no tener falta de información o incluso de funcionamiento. El operador Decision Tree que estaba dentro del Cross Vaidation podía ser un poco delicado al leer los datos, por eso se debía intentar no sobrecargar para que el Perfomance diera resultados confiables.

Con la siguiente tarea que se asignó, tuvo como primer reto el lograr leer los datos ya que venían de manera no legible para el sistema de RapidMiner, al adaptar estos el programa reproduce los datos de manera adecuada al igual que simplifica muchos posibles futuro inconvenientes. Luego de logar limpiar los datos a gusto, se continua generando el modelo, este tenía como enfoque predecir el volumen de nuevos productos que se van a introducir a la tienda. El primer modelo debe hacerle el training a la información. Parte fundamental de esto es elegir los atributos relevantes para la predicción por medio de un análisis de correlación, los que tengan mayor relación con el atributo a elegir es lo recomendado. El siguiente paso es buscar el operador de predicción que mejor se adapte al enfoque que se busca, el GBT mostro índices bajos de confianza. El SVM, para poderlo usar se debió eliminar el ProductType ya que este no lee datos polinomiales, las predicciones tampoco mostraron los índices buscados, por lo tanto se terminó usando K-NN con 3 K dentro del Cross Validation, ya que este dio los mejores número y el resultado se determinó confiable, para medir la confianza se usó Perfomance Regression. El uso de pocos atributos que realmente contribuyeran eran necesarios para no tener Overfitting. Los datos deben estar bien normalizados. Luego se crea un segundo modelo casi idéntico para predecir los datos de los nuevos productos. Al dar conclusiones y al buscarlas se debe tener claro el norte a seguir. Se debe revisar que la información revelada tenga congruencia con lo que se busca responder.

**Como cambiarias aspectos del proceso que no hayan funcionado bien?**

Analizando y probando, se debe utilizar diferentes combinaciones para poder dar con el resultado que se espera, cada cambio debe ser con sentido, intentando dar con una respuesta o predicción establecida. Después de cada cambio se debe revisar si este contribuyo con el fin. De no ser así se debe seguir cambiando hasta que se vean resultados confiables.   
Informándose sobre el objetivo y las herramientas con las que se cuentan.  
Preguntando cada vez que se tiene dudas sobre algún resultado obtenido, sin dejar de lado algunos modelos de los cuales se tienen dudas a expertos en temas.

**Qué funcionó bien utilizando RapidMiner y que no?**

Por lo general la experiencia con RapidMiner fue bastante interesante, bastante sencillo de usar, con una gran variedad de tutoriales y de información general sobre los operadores.  
Este no funciona bien si los datos no se encuentran limpios.  
Por momentos el programa funcionaba de manera lenta, dependiendo que tan pesado sea el modelo.  
Aunque algunos operadores no son muy configurables si estos tienen los datos de manera clara se podrán ver resultados confiables.

**Que consejos que le darías a otro especialista en análisis de datos realizando un proyecto similar?**

Prediciendo éxito de ventas cruzada

Las lecciones generalmente son bastante importantes en cualquier ámbito en que se desarrollen, pero en este caso nos enfocamos en lecciones prácticas para el mundo real de los negocios, los resultados tendrás que ser aplicados y de no evaluarse bien re podrían generar pérdidas significativas.  
El siguiente texto enseñara como resaltar tendencias de compras de los clientes, mostrar números relevantes que generen cambios significativos y realizar predicciones de manera correcta a productos a integrar.

Para iniciar se deben cargar los datos en RapidMiner, agregarle el operador “Select Attribute”, convertir los atributos del In.Store en binomiales por medio de “Numeric to Binomial”, hacer lo mismo a el atributo Region con “Numeric to Polinomial”, después discretizar un subset de atributos en 3 segmentos (Amount, Items y Age) por medio de “Discretize by Bining”, luego con “Set Role generarle un label a la Región, a continuación se introduce el operador “Cross Validation” dentro de el se debe agregar en el training “Decision Tree” y en el testing “Apply Model” con “Perfomance”. Con eso se van a obtener las predicciones sin embargo se debe de estar analizando los datos con cada operador que se agrega para responder las dudas de la compañía.  
Es importante tomar en cuenta que todos los puertos de entrada y salida estén correctamente conectados.  
Los datos se deben de analizar en relación con lo que se desea encontrar.  
No sobre cargar de segmentos la predicción.  
  
Para la segunda tarea se cargan los datos de los productos existentes luego de haberlos limpiados correctamente para su respectiva lectura de parte del programa, se agrega “Select Attributes”, se seleccionan los atributos relevantes para la predicción, anteriormente se tuvo que hacer un cuadro de correlación para determinar los atributos adecuados, pero en este caso nos brincamos el proceso y seleccionamos: PositiveServiceReview, ProductType, ProductNumber, Volume, x5StarReview y x4StarReview. Luego se deben poner 3 operadores de “Set Role”, uno para ponerle un ID al ProductType y el otro ponerle un ID a ProductNumber, al tercero se le selecciona un Label a Volume, antes de poner el “Cross Validation” se deben haber normalizado los números con “Normalize”, una vez agregado ambos dentro del cross, en Training se agrega “K-NN” con 3 K, y en el testing “Apply Model” con “Perfomance Regression”, a este se le deben seleccionar los índices de confianza que se crean necesarios, por último se almacenan los datos con “Store”  
Se inicia un modelo nuevo dentro del mismo proceso, el primer debe ser completamente deshabilitado luego de almacenarlo, el segundo inicia con la información de los productos nuevos que se buscan predecir, con los mismos operadores a excepción del “Cross Validation” que se reemplaza por “Apply Model” y se le hace una conexión doble desde el proceso que se viene haciendo y la recuperación de los datos almacenados del primer modelo. Luego de que todo esto esté correctamente conectado se deberá ver las predicciones de los nuevos ítems.  
Los datos entre más limpios estén, mejor serán leídos.  
Hay que mantener el proceso sencillo y saber seleccionar solo los datos relevantes.  
Mantenerse siempre en dirección de lograr los objetivos