**qwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmrtyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnmqwertyuiopasdfghjklzxcvbnm**

|  |
| --- |
| Reporte de Lecciones  Alert! Data Analytics  1/16/2019  Esteban Vargas |

**Reporte de Lecciones**

**Galaxy**

Para desarrollar la predicción al set de datos del celular Galaxy, se utilizó el Gradient Booster Model, GBM como modelo. EL set de datos para entrenar el modelo, contenía todos los atributos que originalmente traía el documento, ya que así fue como se presentó la mejor exactitud.

Se probaron con diferentes modelos como C5.0, Random Forest, KKNN, Ranger y GBM, pero el que dio mejores número de desempeño fue el GBM y por ende fue elegido para el modelo final.

Primeramente se realizó un modelo Out of the Box, con todos los atributos para tener una base con los desempeños, luego de probar diferentes set de datos con variaciones en los atributos se llego a la conclusión que el set de datos original daba mejores resultados que los otros que se intentaron como menos atributos.

A continuación se presenta un cuadro con los resultados de los modelos utilizados.



**Iphone**

Para generar el modelo predictivo del Iphone se utilizó el Random Forest RF; al igual que para el Galaxy, se entrenó el modelo con todos los atributos en el set de datos ya que así se dieron los mejores resultado de exactitud.

Accuracy y Kappa mostraron mejores números utilizando el RF.

Al igual que para el Iphone, no se eliminó ningún atributo ya que eliminarlos no mejoro los numero de exactitud.

A continuación se presenta un cuadro con los resultados de los modelos utilizados.



**Proceso**

Por lo general durante todo el proceso de modelaje no se encontró con tantos inconvenientes como de costumbre, los datos se encontraban en casi perfectas condiciones para ser modelados, solo hubo que cambiar los valores a numéricos y el atributo a predecir se convirtió en un factor.  
Para dar con un mejor resultado, se crearon varios set de datos utilizando en cada uno una combinación diferentes de atributos para determinar cuáles eran los más importantes para dar con niveles altos de exactitud. Luego de varias pruebas se concluyó que el modelo era más eficiente daba con todos los atributos de igual forma fue un poco difícil determinar el mejor ya que todos tenían números muy similares, para esto se realizaron cuadros comparativos.  
Números parecidos también dieron en los diferentes modelos que se probaron, algunos no encajan bien con el tipo de dato y daban resultados bajos por lo tanto no se agregaron al reporte, en el reporte solo de dejaron los 4 mejores modelos.  
Durante el proceso es importante llevar un control y orden de lo que se está haciendo, saber probar cada modelo y set de datos sin destruir el código o confundirse con otro.   
Algunas veces los códigos tienen ciertos errores que hace que el resultado que se da no es el correcto.

**Recomendaciones**

Para futuros proyectos se debería intentar de extraer datos de una cantidad mayor de páginas webs, aunque eso haga el proceso más lento pero tal vez se puedan encontrar patrones distintos y más marcados.  
Utilizar palabras más específicas para sustraer mejor la información relevante.  
Agregarle más modelos al código actual para poder probar con más opciones a elegir.  
Se podría también en lugar de utilizar 4 categorías de sentimientos, podrían usar solamente 3, “positivo”, “neutral” y “negativo”.