

Comprensión y predicción de la calidad fotográfica móvil

Yusef Ahsini Ouariaghli
Mikel Baraza Vidal
Pablo Díaz Masa Valencia
David Canet Llorca
Bamba Mohamed Salek



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Proyecto II, Integración y Preparación de Datos
Grado en Ciencia de Datos

Índice

1. Alcance del proyecto
 - 1.1. Objetivo del proyecto
 - 1.2. Utilidad del estudio
2. Configuración del proyecto
 - 2.1. Fuentes de datos
 - 2.2. Integración y transformación de los datos
 - 2.2.1. Integración de los datos
 - 2.2.2. Transformación de los datos
3. Resultados obtenidos
 - 3.1. Calidad fotográfica móvil y su relación con los componentes.
 - 3.1.1. Clustering
 - 3.1.2. Predicción de calidad fotográfica
 - 3.1.2.1. Calidad fotográfica final
 - 3.1.2.2. Calidad foto nocturna
 - 3.2. Estudio de la percepción subjetiva de la fotografía móvil.
 - 3.2.1. Análisis de sentimientos
 - 3.2.2. Text-Mining
 - 3.2.3. Predicción de sentimientos
 - 3.2.4. Conclusión
 - 3.3. Privacidad y legalidad de los datos
4. Lecciones aprendidas para mis futuros proyectos en Ciencia de Datos.

1. Alcance del proyecto

“Comprensión y predicción de la calidad fotográfica móvil”

Para el proyecto nos hemos planteado estudiar la calidad de las cámaras móviles, de modo que podamos crear una guía para entender mejor este campo y así poder ofrecerlo tanto al usuario para que pueda realizar una mejor compra, como a las empresas de modo que puedan crear un mejor producto.

Para ello, hemos realizado estudios de una lista de 102 móviles, en primer lugar, de las especificaciones técnicas de las cámaras, complementándolo con un conocido test fotográfico puramente objetivo, y a su mismo tiempo la opinión y experiencia de la gente al respecto de dichos móviles, recopiladas a través de reseñas realizadas en internet.

1.1. Objetivos del proyecto

Objetivo 1: Realizar un estudio de la fotografía móvil desde una perspectiva técnica.

- Comprender qué componentes y características físicas y de software afectan positiva y negativamente a la calidad fotográfica del dispositivo.
- Predecir la calidad fotográfica objetiva de un dispositivo en función de su hardware y software.

Objetivo 2: Comprender la percepción subjetiva de la calidad fotográfica a través de las reseñas.

- Entender cómo afectan las distintas variables (características) a la percepción de la calidad de las fotos en los usuarios finales.
- Confirmar o descartar nuestras hipótesis respecto a las relaciones entre calidad técnica, calidad percibida, marca y precio.

1.2. Utilidad del estudio

Con este estudio tratamos de ayudar a responder la pregunta que tanta gente se realiza:

“¿Qué dispositivo móvil debo comprar según mis intereses?”

Ante la inmensa cantidad de teléfonos móviles a la venta, en múltiples ocasiones se hace imposible llevar a cabo una única elección. Además de ello los dispositivos ofrecen diferentes propiedades, sobre todo tratamos de ayudar a ese público que trata de llevar a cabo la compra del dispositivo girando entorno a la cámara, facilitando la cantidad de opciones y aportando información para encontrar una elección tanto económica como de calidad.

También la otra utilidad de nuestro estudio es que los fabricantes de smartphones, mediante los hallazgos encontrados en este estudio creen mejores dispositivos.

2. Configuración del proyecto

2.1. Fuentes de datos.¹

Durante la búsqueda de datos para nuestro proyecto, no hemos encontrado ningún data-set con los datos que estábamos buscando y por tanto hemos recurrido a técnicas de web-scraping para recolectarlo de distintas webs usando scripts Python. Los datos relacionados con las especificaciones de los smartphones han sido obtenidos de KiMovil, los datos del test fotográfico provienen de DXOmark y los comentarios de smartphones de Google Shopping.

2.2. Integración y transformación de los datos.²

Para realizar nuestro estudio, hemos obtenido dos ficheros de datos, uno que recoge las características y propiedades de todos los móviles que queremos analizar y otro que recoge los comentarios y valoraciones de miles de usuarios. El fichero de las características y propiedades lo hemos obtenido integrando la base de datos obtenida de KiMovil con los resultados del test fotográfico obtenidos de DXOmark. Por otro lado, el fichero de los comentarios lo hemos conseguido integrando el fichero de reseñas obtenido de Google Shopping con el léxico Afinn con los que realizamos el análisis de sentimientos y clasificamos cada comentario.

3. Resultados obtenidos

3.1. Calidad fotográfica móvil y su relación con los componentes.

En esta parte del estudio tratamos de responder a la pregunta: *"¿Qué hace que una cámara sea buena?"*.

Esta pregunta va dirigida a todos aquellos que en algún momento han dudado sobre que teléfono móvil comprar ante la cantidad abrumadora de dispositivos que hace imposible distinguir entre aquellos que responden mejor a esta pregunta. La compra de un teléfono está fuertemente guiada por los intereses del usuario. Es por esto que nos centramos en aquellos usuarios que desean saber cuáles son esos componentes o especificaciones del teléfono que supondrán una mejora fotográfica.

3.1.1. Clustering.

¹ Más información sobre los datos utilizados en el proyecto en el Anexo 1

² Más información sobre integración y transformación en el Anexo 2

Vamos a realizar un clustering con el objetivo de agrupar los móviles con características similares y así poder determinar que patrones siguen. Las librerías que vamos a usar para realizar el clustering son **cluster**, **FactoMineR**, **factoextra**, **NbClust**, **clValid** y **fpc**.

Al tener en nuestros datos un porcentaje similar de variables categóricas y numéricas, decidimos utilizar la distancia de Gower. Al tratarse de una distancia de similitud, permite ver qué tan distintos son dos registros y funciona con registros mixtos. Vamos a utilizar dos métodos de clustering jerárquicos (el de Ward y la media) y K-medoides. Antes de aplicar los algoritmos comprobamos que existe tendencia de agrupación en nuestros datos mediante un gráfico de correlación, con el que deducimos que existen dos grupos grandes de teléfonos similares y varios de menor tamaño.

Para el método de Ward vamos a seleccionar 6 clusters. Sin embargo, para el método de la media escogeremos 5, ya que son los puntos donde obtenemos un buen coeficiente de silhouette. Podemos examinar cómo los tamaños de los cluster de Ward son bastantes equitativos, mientras que los clusters de la media son desiguales, debido a que hay un cluster con 68 teléfonos y otro con solo uno, siendo éste un Sony Xperia 1, uno de los móviles con menos características.

Continuando con el clustering, vamos a realizar el algoritmo de K-medoides. Para obtener el número ideal de clusters hacemos uso del coeficiente de silhouette, que en este caso es 8.

Para seleccionar el mejor método, comparamos los tres clustering que hemos realizado calculando el coeficiente de silhouette para cada uno de los individuos, y nos quedamos con el **método de Ward**, ya que es el algoritmo que tiene menos individuos mal colocados. Ahora haciendo uso de un T-SNE para comprimir nuestros datos a dos dimensiones, observamos que los clústeres generados están bien diferenciados (Figura 1).

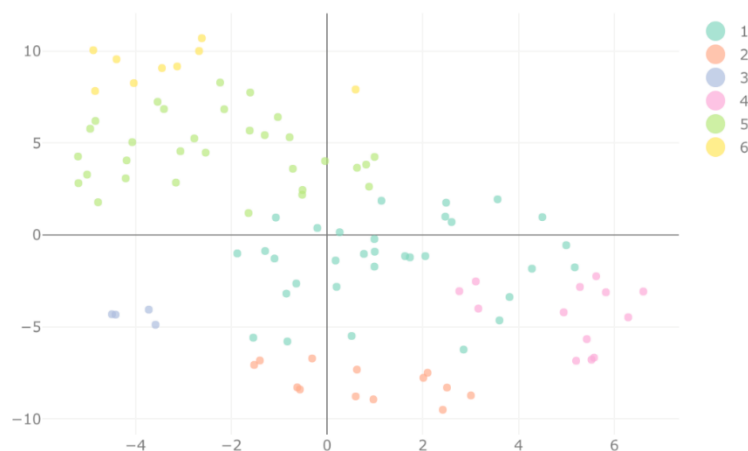


Figura 1. T-SNE con los clústeres generados coloreados.

Observando los perfiles de nuestros clusters podemos distinguir que los grupos generados están segregados por categoría de producto, separados en grupos de dos en gama alta, media y baja.

Dentro de la gama alta podemos distinguir a los móviles con componentes similares, pero con la diferencia de que los que mejores resultados obtienen dentro de la gama alta son aquellos que usan sensores con menor resolución y apertura (cuanto más baja la apertura más luz entra al sensor). También apreciamos que todos los dispositivos de este grupo poseen estabilización óptica.

En la gama media observamos que los grupos que nos genera son similares a nivel de resultado fotográfico, que se han separado en un grupo donde los dispositivos poseen más de una cámara y los que solo tienen una. La mayoría de los dispositivos en este grupo poseen estabilización óptica.

Finalmente, en los grupos de móviles de gama baja se observa que los grupos con peores fotos y resultados son aquellos con más cámaras, ya que el coste de estas cámaras lastra la calidad de componentes importantes para la calidad fotográfica como el procesador.

La conclusión de este apartado es que en los móviles más baratos es importante evitar comprar los que mayor número de cámaras disponen y valorar más la calidad de otros componentes. En la gama media el número de cámaras no afecta a la calidad y añaden funcionalidades, por lo tanto, se pueden elegir en función de si interesan estas funciones añadidas. Por último, en la gama alta es recomendable evitar los sensores con mayor número de megapíxeles y apertura si lo que se busca es la mejor calidad posible. Cabe destacar que es importante fijarse en si el dispositivo dispone de estabilización óptica para obtener los mejores resultados.

3.1.2. Predicción de calidad fotográfica.

En este apartado del proyecto vamos a intentar comprender como interactúan nuestras especificaciones con el resultado fotográfico de los móviles. La medida que utilizamos para medir si una cámara es mejor o peor es la puntuación obtenida en el test de DXOMark. Por ello, mediante un modelo de regresión, la explicaremos en función de las especificaciones técnicas de los móviles.

3.1.2.1. Calidad Fotográfica Final.

En este caso estudiamos la calidad fotográfica (*final_score*), siendo esta la nota final obtenida según la puntuación de cada una de las especificaciones de cámara. *final_score* se usa como variable predicha y el resto de los valores numéricos de las especificaciones técnicas como variables predictoras. No tiene sentido usar las variables de valoración de los aspectos de la cámara ya que la *final_score* se extrae a partir de estas y ya están relacionadas.

El modelo se ajusta a partir de la muestra de entrenamiento y finalmente se usa para predecir las respuestas de la muestra de testeo. A continuación, pasamos a describir los resultados y trabajo de este modelo.

Gracias a la función base de R **lm** generamos modelo de regresión con todas esas variables que usaremos como predictoras, es entonces donde se lleva a cabo la selección de esos valores utilizados para el **train** y seleccionamos todas aquellas variables que resultan significativas en el modelo mediante el algoritmo *backward stepwise selection* por el cual obtenemos el mejor modelo posible (Figura 2).

$$\begin{aligned} final_score = & 113,7787 + 0,00003073314 * Puntuacion_CPU + 0,006475693 \\ & * Precio - 2,361076 * Nanometros + 0,1538123 \\ & * Resolucion_2 - 6,979196 * Tama\tilde{n}o_pixel_2 \end{aligned}$$

Figura 2. Modelo que explica la calidad fotográfica de un smartphone.

Debemos tener en cuenta que no podemos tener en cuenta valores como *tamaño_pixel_2* o *resolución_3* ya que son variables que poseen un número muy elevado de valores nulos y esto hace que no sean verdaderamente importantes para nuestro estudio. Estos nulos aparecen ya que gran parte de los dispositivos no poseen segunda o tercera cámara.

Con la ecuación resultante generamos el nuevo modelo que será utilizado para comparar la predicción realizada con los valores extraídos inicialmente con el *test*. Al visualizar el R^2 obtenemos un valor de 0.573. Esto tiene sentido debido al número reducido de variables utilizadas e indica que sí representan gran parte de la variabilidad de la variable respuesta.

Observando el modelo ya podemos inferir dos cosas: la importancia de la potencia de procesamiento en la calidad fotográfica de los smartphones actuales y la necesidad de una buena cámara secundaria, evitando así los sensores de peor calidad.

3.1.2.2. Calidad Foto Nocturna.

A continuación, estudiamos la fotografía nocturna, ya que consideramos que al ser uno de los valores más importantes dentro de la cámara resulta interesante para realizar el estudio. Utilizamos como variables predictoras las mismas utilizadas anteriormente. Después de aplicar el algoritmo de *backward stepwise selection* obtenemos nuestro modelo (Figura 3).

$$\begin{aligned} Photo.Night = & 16,22599 + 0,00003307770 * Puntuacion_CPU + 0,8566673 \\ & * Ram - 2,397742 * Nanometros + 13,38485 * Tama\tilde{n}o_pixel_1 \\ & + 0,1900855 * Resolucion_2 \end{aligned}$$

Figura 3. Modelo que explica la calidad fotográfica nocturna de un smartphone.

Para la fotografía nocturna obtenemos que se repite lo encontrado en el modelo anterior, añadiendo la importancia de tener un tamaño de pixel mayor en el sensor principal.

3.2. Estudio de la percepción subjetiva de la fotografía móvil.

En la parte anterior de este estudio hemos tratado de comprender qué componentes hacen que una cámara y su rendimiento sean objetivamente buenos, mientras que en esta lo que vamos a buscar es descubrir qué contribuye a que un usuario perciba una buena experiencia y considere una cámara buena. Para ello, en este apartado nos centramos en comprender cómo reaccionan los usuarios a los distintos smartphones de nuestra base de datos y qué características impactan más en su percepción de calidad. Con esta intención, hacemos uso de nuestra base de datos de comentarios sobre teléfonos inteligentes con su valoración asociada.

3.2.1. Análisis de sentimientos.

En primer lugar, vamos a realizar un análisis de sentimientos de todos los comentarios usando las librerías **tidytext** y **dpplr**. Además, repitiendo el mismo proceso y usando un léxico fotográfico propio, obtenemos cuántas palabras relacionadas con la fotografía contiene cada comentario para poder filtrarlos posteriormente (Figura 4).

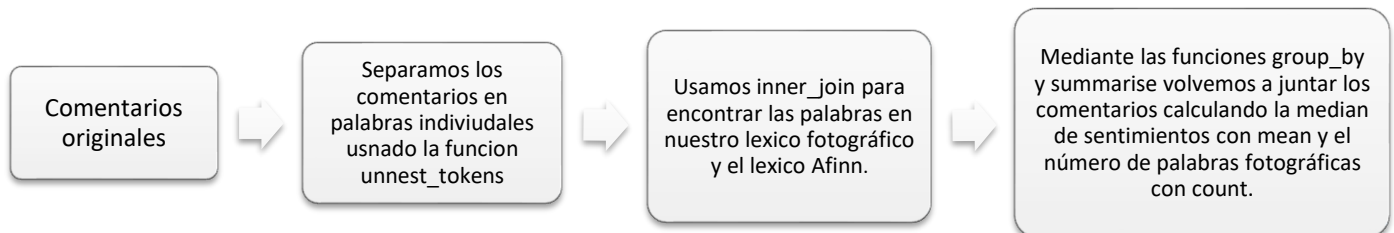


Figura 4. Análisis de sentimientos.

Una vez disponemos de una puntuación media de sentimiento para cada uno de nuestros 80.000 comentarios, centramos las valoraciones asociadas y la puntuación asociada a los comentarios en un rango comprendido entre -1 y 1. Para este análisis concreto y con el objetivo de trabajar con comentarios relacionados con la fotografía, filtramos los comentarios y conservamos únicamente aquellos con aquellos que contienen cuatro o más palabras de nuestro léxico fotográfico. Con este filtro aparecen correlaciones esperadas, como son la correlación negativa entre los sentimientos y valoraciones con la resolución del sensor principal, similar a la que existe entre esta resolución y la puntuación final del test fotográfico.

3.2.2. Text-Mining.

Gracias al filtro aplicado a los comentarios, ahora disponemos de una colección de comentarios con información relevante y relacionada con el

tema de estudio: la calidad fotográfica. En este apartado nos dedicamos a extraer conocimiento de nuestros comentarios mediante algunas técnicas de minado de textos.

Para entender qué hechos afectan más a los usuarios tanto positiva como negativamente haremos dos divisiones de nuestros comentarios originales. La primera de ellas es separar los comentarios en grupos por la valoración asociada al comentario (entre 1 y 5 estrellas) y la segunda separarlos en tres grupos por puntuación de sentimiento (-0.5-0, 0-0.5, 0.5-1).

Una vez creados los grupos, con el objetivo de analizar el contenido textual de nuestros comentarios, aplicamos un proceso de unificación y limpieza que realizamos usando la librería de R **tm**, diseñada para text-mining. Usando la función base de R *paste*, unificamos todos los comentarios de nuestro dataframe en una única cadena de texto. Después, haciendo uso de las diferentes funciones de **tm**, eliminamos los números y signos de puntuación, pasamos el texto a minúsculas y eliminamos las stopwords (palabras muy comunes que no aportan ninguna información distintiva).

Con el texto limpio, buscaremos en el texto la frecuencia de cada palabra así como la de los 3-gramas (grupo de tres palabras) y los ordenamos en orden descendente. Gracias a esto podemos visualizar qué quejas y descontentos son más recurrentes entre los usuarios.

Después de realizar esto para los dos grupos de comentarios, podemos llegar a la primera conclusión: las cámaras secundarias afectan marcadamente a la percepción de calidad de los usuarios (Figura 5). Este efecto se aprecia tanto en los comentarios con peor puntuación de sentimiento como en los comentarios con peor valoración otorgada.

Además, lo ilustramos generando un wordtree (usando la librería **wordtreer**) sobre la palabra lens utilizando nuestros comentarios con menor puntuación de sentimiento. El resultado es que las palabras que la acompañan suelen tener connotaciones negativas (Figura 6), mientras que en los comentarios con puntuación de sentimiento media y alta esto no ocurre y las palabras que acompañan a la palabra lens son mayoritariamente positivas (Figura 7).

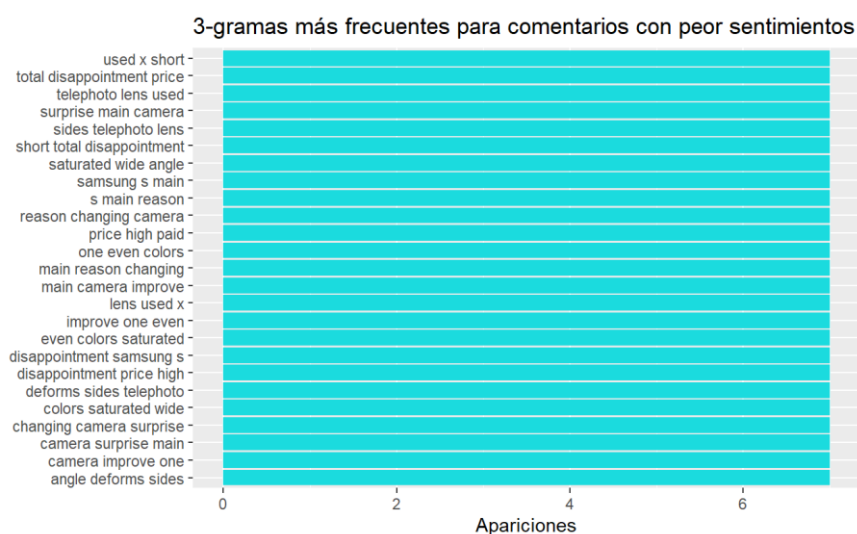


Figura 5. 3-gramas más frecuentes para comentarios con peor sentimientos.

Este resultado cobra sentido cuando tenemos en cuenta que estas lentes secundarias suelen ser un recurrente reclamo de marketing que genera grandes expectativas a los clientes. Cuando el usuario hace uso de estas lentes secundarias, se da cuenta de que usan sensores de peor calidad que las camaras principales.

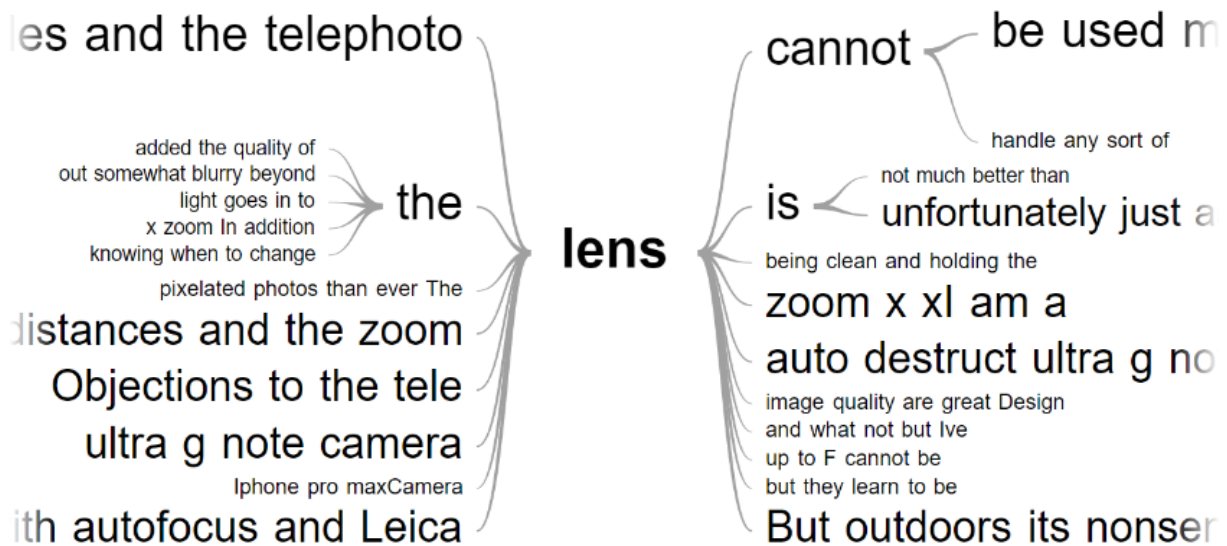


Figura 6. Wordtree con la palabra lens como centro para comentarios negativos.

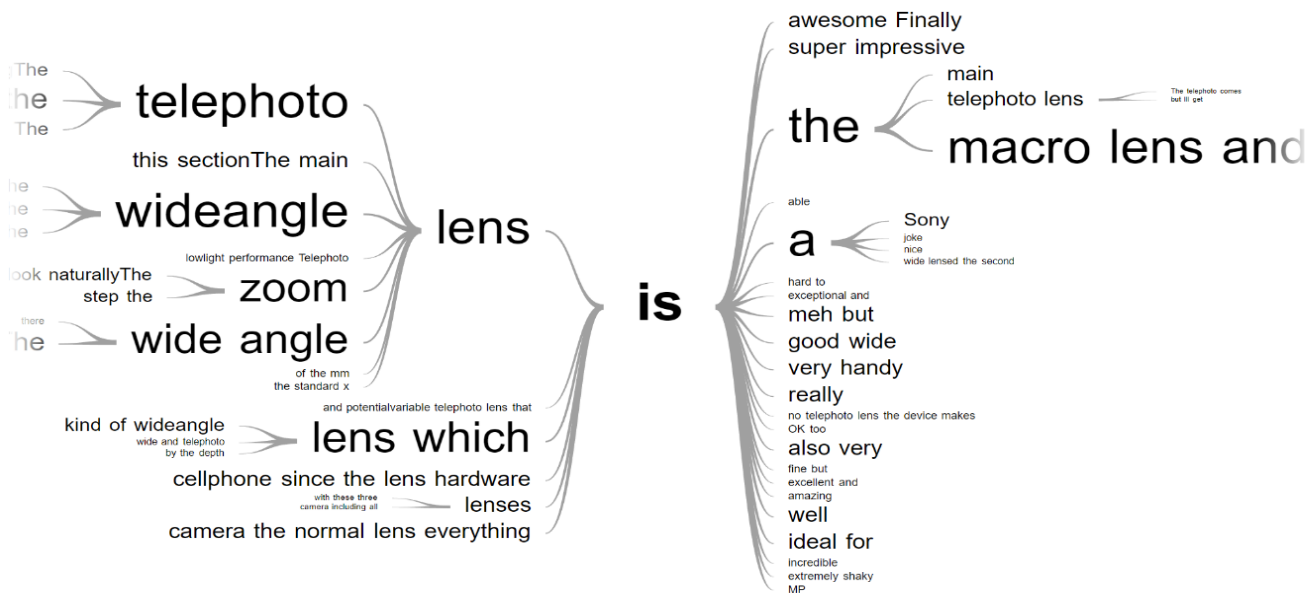


Figura 7 Wordtree con la palabra is como centro para comentarios con sentimientos intermedio y positivos.

3.2.3. Predicción de sentimientos.

Para completar nuestra comprensión de la actividad de los usuarios, vamos a predecir su comportamiento mediante un modelo de regresión lineal que

explique la puntuación de sentimiento que inspira la cámara de un determinado smartphone en función de sus componentes y sus puntuaciones del test de DXOMark. Para realizar dicho modelo, integramos nuestras dos bases de datos: por un lado, la de sentimientos; por otro, la de especificaciones y puntuaciones del test fotográfico (Figura 8).

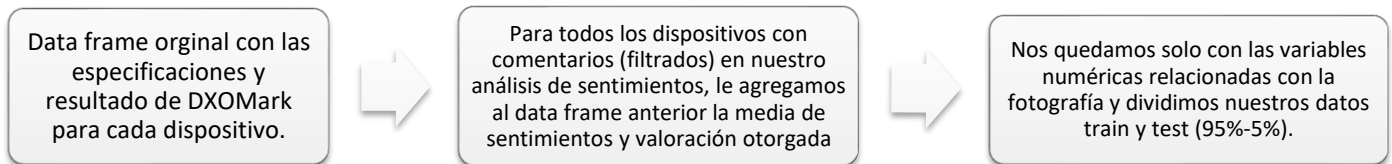


Figura 8. Integración de datos para modelo de regresión lineal.

Con estos datos, y haciendo uso de la función base de R **lm** generamos un modelo de regresión con todas las variables disponibles en nuestros datos integrados. Después, haciendo uso de la función **step**, aplicaremos un algoritmo de *backward stepwise selection* con el que eliminar las variables hasta obtener el mejor modelo posible. Después de aplicar el algoritmo obtenemos como resultado nuestro primer modelo valido (Figura 9).

$$\begin{aligned}
 \text{sentimiento} = & 0.380477794 + 0.004976398 * \text{Photo.Bokeh} - 0.004045087 \\
 & * \text{Photo.Color} + 0.005244642 * \text{Photo.Preview} - 0.004948504 \\
 & * \text{Video.Color} + 0.001581797 * \text{Video.Exposure} - 0.003068235 \\
 & * \text{Video.Noise} + 0.001969849 * \text{Video.Stabilization} + 0.001285803 \\
 & * \text{Video.Texture} - 0.001444796 * \text{Zoom} - 0.000000119 \\
 & * \text{Puntuacion_CPU} + 0.000676541 * \text{Peso} - 0.000101353 * \text{Precio} \\
 & + 0.071876135 * \text{Estabilizacion_optica}
 \end{aligned}$$

Figura 9. Primer modelo de regresión lineal que explica el sentimiento medio generado por una cámara.

Observando los residuos de nuestro primer modelo comparado con las variables vemos algunos patrones que indican no linealidad en las variables Zoom y Video.Stabilization, por lo que generaremos un segundo modelo con éstas al cuadrado (Figura 10).

$$\begin{aligned}
 \text{sentimiento} = & 0.446486460 + 0.004646341 * \text{Photo.Bokeh} - 0.004273803 \\
 & * \text{Photo.Color} + 0.005306185 * \text{Photo.Preview} - 0.004909488 \\
 & * \text{Video.Color} + 0.001730990 * \text{Video.Exposure} - 0.003063076 \\
 & * \text{Video.Noise} + 0.000010729 * \text{Video.Stabilization}^2 + 0.001266381 \\
 & * \text{Video.Texture} - 0.000010649 * \text{Zoom}^2 - 0.000000118 \\
 & * \text{Puntuacion_CPU} + 0.000679512 * \text{Peso} - 0.000101910 * \text{Precio} \\
 & + 0.071242728 * \text{Estabilizacion_optica}
 \end{aligned}$$

Figura 10. Segundo modelo de regresión lineal que explica el sentimiento medio generado por una cámara.

Para escoger qué modelo explica mejor la puntuación de sentimiento elegiremos el que menor error cuadrático medio tenga y mayor r^2 , que en este caso es el segundo modelo con 0.02660253 de RMSE y 0.58794063 de

r2. Basándonos en el RMSE, podemos afirmar que nuestro modelo es bastante preciso, ya que las predicciones del mismo sólo se desvían 0.026 de la realidad. El rango de sentimientos que genera el modelo estará entre -0.1 y 0.4 al ser entrenado con datos en dicho intervalo.

3.2.4. Conclusión.

Después de todos los análisis anteriores, ya entendemos qué factores rigen la percepción de los usuarios de la calidad fotográfica. Usando el análisis de sentimientos como medida de si esta percepción es buena o mala, podemos sacar varias conclusiones.

La primera de ellas es el impacto de las cámaras secundarias en la percepción de calidad, algo que extraemos tanto del apartado de text-mining, con los principales 3-gramas como en el wordtree, como en el modelo de regresión, donde la puntuación de Zoom tiene coeficiente negativo (esta puntuación valora más el número de distintas cámaras que la calidad y utilidad de las mismas). Estas cámaras generan unas expectativas al usuario que después no se cumplen, ya que al momento de usar la cámara del móvil el usuario no distingue cuál de ellas utiliza, pero al ser estas cámaras secundarias de baja calidad debido al uso de peores sensores, las fotos hechas con ellas no gozan de una calidad tan alta como la de las fotos hechas con la cámara principal.

También podemos estimar qué variables y en qué medida afectan significativamente a la percepción de sentimientos gracias al modelo de regresión.

Las variables que afectan positivamente a la percepción de calidad del usuario son Photo.Bokeh, Photo.Preview, Video.Exposure, Video.Stabilization, Video.Texture, Estabilizacion_optica y peso. Las variables Photo Bokeh, Photo Preview, Video Exposure, Video Stabilization y Video Texture son simplemente puntuaciones de diferentes aspectos técnicos que encontramos en los test de DXOmark. Sin embargo, la estabilización óptica es una especificación ajena a este test que también colabora positivamente en la experiencia del usuario, pues las fotos suelen salir más nítidas a pesar del movimiento del móvil al tomarlas. Finalmente, el peso afecta positivamente a la puntuación de sentimiento porque un mayor número de componentes fotográficos especializados conlleva un peso mayor.

Por otra parte, las variables que afectan negativamente son Photo.Color, Video.Color, Video.Noise, Zoom, Puntuacion_CPU y Precio. Las variables Photo Color, Video Color, Video Noise y Zoom vuelven a ser simplemente puntuaciones técnicas de DXOmark. No obstante, cabe señalar que el Zoom tiene el efecto ya explicado anteriormente de elevar insatisfactoriamente las expectativas del usuario. También las variables Puntuacion CPU (que está íntimamente ligada al precio, pues es uno de los componentes más caros y distintivos del móvil) y Precio afectan negativamente a la percepción del usuario por esta razón.

3.3. Privacidad y legalidad de los datos.

Nuestros datos son especificaciones de dispositivos móviles y comentarios anónimos y públicos en Internet, por lo que no están ligados a individuos concretos y en principio no resulta difícil respetar la ley y la privacidad. Es por esto que no hemos tenido problemas de este tipo a la hora de obtener los datos.

Si bien DXOmark no hace públicos los métodos y fórmulas de sus tests, las puntuaciones de éstos, que son los primeros datos que conforman el fichero de especificaciones, son fácilmente accesibles en su página web. La dificultad reside más bien en extraer las cifras de los objetos JavaScript incrustados en la página. Es por esto que, respecto a la privacidad, al ser datos públicos y no personales no hay problema en sacarlos de la página web.

La obtención de las características técnicas de los diferentes dispositivos, que conforman la segunda parte del fichero de especificaciones, tampoco acarrea problemas de índole legal, pues son datos que están disponibles en la página web de Kimovil sin necesidad de suscribirse ni pedir ningún tipo de permiso. Tampoco afectan a empresas ni individuos, así que se pueden utilizar.

Por último, los comentarios que nos sirven de punto inicial para realizar el análisis de sentimientos los hemos sacado de Google Shop. Al ser comentarios públicos que no contienen información identificativa ni sensible, tampoco existen aquí problemas de privacidad. Además, no hemos recopilado nombres de usuario, nombres de cuenta ni correos electrónicos, pues no son relevantes para nuestro análisis.

4. Lecciones aprendidas para mis futuros proyectos en Ciencia de Datos.

Creemos que no podemos dudar a la hora de calificar este proyecto como satisfactorio, pues, además de haber obtenido buenos resultados, hemos aprendido y perfeccionado diferentes métodos de análisis.

Hemos llevado a cabo varios análisis que no conocíamos antes, como pueden ser el clustering con variables tanto numéricas como categóricas, el análisis de sentimientos a partir de un léxico y el text-mining.

Además, si bien hemos usado Python (Jupyter Notebook) para webscrapping, hemos explotado bastante más el lenguaje R y las opciones que ofrece RStudio: esto es algo importante y útil de cara a futuros proyectos y asignaturas.

Por último, varias aplicaciones han resultado útiles para comunicarnos entre los miembros del equipo (Teams), llevar un registro histórico actualizado de los conjuntos de datos (Discord), elaborar documentos (Word) y exponer resultados (PowerPoint).

ANEXOS

Anexo 1

En este anexo se encuentra toda la información relativa a la captura de datos y el código necesario para obtenerlos.

Archivos: **Captura de datos.pdf**, `get_dxomark.ipynb`, `get_kimovil.ipynb`, `get_reviews.ipynb`

Anexo 2

Toda la información sobre la integración y análisis exploratorio de nuestras bases de datos.

Archivos: **Integración y preparación de datos.pdf**, `Limpieza de datos.html`

Anexo 3

El código R del estudio y mediante el cual hemos aplicado las distintas técnicas estadísticas para obtener las conclusiones.

Archivos: **Clustering cámaras.html**, **Regresión fotografía.html**, **Análisis datos y sentimientos.html**.

Anexo 4

Archivos de datos usados en el proyecto, contienen la información de los dispositivos y los comentarios de los usuarios.

Archivos: **specs.csv**, **english_reviews.csv**.