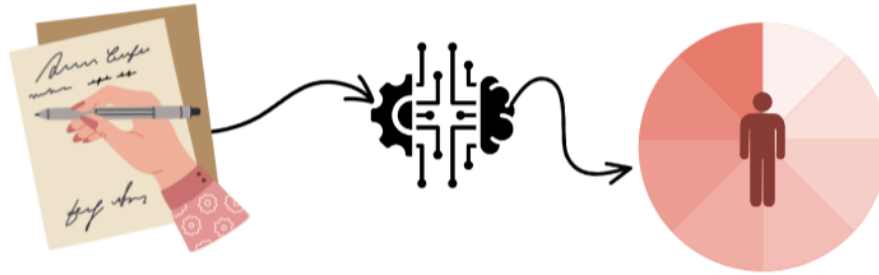


Aprendizaje de Máquinas

Análisis Grafológico



Sheyla Leyva Sánchez
Alejandra Monzón Peña
Laura Victoria Riera Pérez
Andry Rosquet Rodríguez
Kevin Talavera Díaz
Javier Villar Alonso

Cuarto año. Ciencias de la Computación.
Facultad de Matemática y Computación, Universidad de La Habana, Cuba

Resumen

La grafología es una técnica que estudia las características psicológicas de las personas a través de su escritura. En algunos centros laborales se utilizan este tipo de análisis para determinar cuando contratar o no a alguien y se emplea también para analizar a criminales y en el proceso de determinar trastornos en personas. Esta es una tarea complicada realizada por expertos, psicólogos y grafólogos, y sus resultados están, numerosas veces, sujetos a la perspectiva de quien los analiza. En este trabajo se propone utilizar algoritmos de Aprendizaje de Máquina, para automáticamente poder predecir características psicológicas de un individuo a partir de un fragmento de su escritura. Con este objetivo se propone la creación de un dataset de imágenes de escritura a mano, así como su uso para entrenar y comparar los modelos para predecir la personalidad basado en el Test de los Cinco Grandes (Big Five). Se implementaron los algoritmos Convolutional Neural Network (CNN), Support Vector Machine (SVM), K-Nearest Neighbours (KNN) y K-Means, estableciendo una comparación de los resultados obtenidos para el análisis grafológico.

1. Grafología

La grafología es el análisis de la escritura manuscrita de un individuo con la intención de determinar rasgos de personalidad. No existe evidencia científica que respalde la efectividad de la grafología, por lo que es considerada una pseudociencia y una práctica científicamente cuestionable.

Tal y como se menciona en el artículo [22] la grafología no se utiliza para predecir el futuro o conocer el pasado de una persona. Su uso está dado como una habilidad para conocer el estado psicológico de la persona en el momento que escribe, incluso se plantea que los resultados de este análisis pueden depender de la edad de la persona o del momento en que redacta; es decir una misma persona puede tener resultados diferentes si se analiza su escritura en dos momentos diferentes de su vida.

Según diversos estudios, si tienes un tamaño de letra grande es probable que seas una persona sociable, valiente y muy segura de ti misma. Por otro lado, si tu letra es mediana, significa que te adaptas fácilmente a las circunstancias y eres abierto al cambio. Por último, si escribes con una letra de tamaño pequeño, tiendes a ser tímido, meticuloso e introvertido.

Si el espacio que incluyes entre las letras es amplio, significa que aprecias tu libertad y que no te gusta sentirte abrumado o rodeado por mucha gente. Si es angosto, no te agrada estar solo y prefieres vivir en

compañía.

En el artículo [15] analizan los rasgos de escritura con el objetivo de crear perfiles para los asesinos, en este artículo explican que rasgos psicológicos se vinculan a: Espaciado de palabras, Espaciado de líneas, Márgenes, Inclinación, Presión del lápiz, Forma de letras i, Forma de letras t, entre otras.

2. Grafología y Aprendizaje Máquina

La combinación de Aprendizaje Máquina y Grafología es abordada en un amplio conjunto de artículos en los que se prueban ideas diversas, e incluso se comparan los resultados obtenidos haciendo uso de diferentes técnicas y combinándolas.

En los artículos [11] y [6] abordan ideas de preprocesado de las imágenes manuscritas haciendo uso del umbral adaptativo Gaussiano y llevando a binario los valores de los píxeles de la imagen atendiendo a un nivel de tolerancia prefijada.

Las Redes Neuronales Convolucionales (CNN) debido a su amplio uso para el trabajo con imágenes, en el artículo [11] se usan las CNN como método propuesto en la arquitectura de una página web que permita determinar la personalidad a partir de la escritura. En otros como [11] se emplea este método en el análisis de firmas, analizando la inclinación y el espaciado entre caracteres, con la limitación de que el dataset utilizado solamente posee 10 firmas de 6 individuos.

De igual modo las CNN conjuntamente con Regresión logística se emplearon en [21] para predecir el comportamiento financiero de una persona, en este caso con un dataset de más de 200 muestras de escrituras de un texto dado y resultados de un cuestionario de acuerdo al Big Five Personality Model.

Otro algoritmo utilizado para el análisis grafológico es SVM, empleado en [6], [20], [14], [8] y [10]. En el caso de [14] emplean el algoritmo sobre firmas de 53 participantes, para obtener un 71 % de efectividad, en otros como [20] y [10] se compara con otros algoritmos, en el primero con ANN y DNN y en el segundo con KNN y árboles de decisión para predecir los resultados del Big Five, en este artículo se utilizó para el SVM un kernel RBF.

También se utilizan: K-Means [17], Fuzzy C-Means [18], ANN [13, 20, 5, 17], Random Forest [12, 4] y KNN [17, 10].

En el artículo [19] realizan un estado del arte sobre grafología y Machine Learning en el cual comparan las investigaciones realizadas en otros artículos respecto a los Features utilizados, las características del dataset, algoritmo empleado y resultados obtenidos.

3. Datasets

Los datasets utilizadas en los estudios de grafología realmente contienen información escasa. En muchos casos son privadas y para acceder a ellas es necesario pagar. En el artículo [12] muestran una tabla comparativa de los datasets disponibles en internet que se han utilizado en otros artículos, en la misma describen cuáles son los tags presentes en cada uno de los datasets, si son públicos o no, en qué artículos se utilizan y de qué año son.

Al analizar este contenido resulta relevante que los tags que están presentes en esos sets de datos son escasos, apenas ponen información, como mano con la que escribe el sujeto, sexo, ciudad en la que vive y poco más.

Se descargaron los datasets de HandwritingDatabase [1] y iam-handwriting-database [2], el primero contenía un conjunto de imágenes de textos manuscritos e información de cada persona como sexo, estado de EEUU donde nació mano con la que escribe y edad, el segundo contenía las imágenes de textos, las mismas

segmentadas por línea, por palabra y por caracteres y tenía información general sobre los textos como ancho de línea, altura de línea, cantidad de palabras por línea, etc; pero ninguna información sobre las personas.

Debido a la dificultad de descargar un dataset y la poca información que ofrecen los datasets disponibles para el problema que deseamos resolver, se decidió elaborar un dataset propia.

4. Creación del dataset

La propuesta es crear nuestro propio dataset, para ello fueron encuestadas 71 personas, a las que se les pidió que escribiesen un texto en español y se les realizó el test de personalidad BigFive [3].

Las hojas manuscritas fueron escaneadas y a las imágenes se les aplicó un preprocesado para llevarlas escala de grises y finalmente a binario con un nivel de tolerancia de 180. Además se realizó una segmentación de los textos en líneas a partir de una difusión de los caracteres utilizando `opencv` de python. Las secciones de imagen que no quedaran del todo segmentadas producto a escrituras muy unidas fueron segmentadas manualmente.

A los test de personalidad se les extrajo automáticamente la información referida a los porcentajes de cada uno de los 5 rasgos distintivos del Big Five, para esto se empleó un código de python que scrapea las páginas web de los resultados del test.

5. Extracción de Features

Para extraer los features de las imágenes del DataSet se utilizaron algoritmos de procesamiento de imágenes. Previamente a las imágenes se les realizó un preprocesado para eliminar ruidos y llevarlas a binario; los features extraídos son: Interlineado (LineSpace), Espaciado (WordSpace), Margen (Margin), Inclonación de la línea (Baseline) e Inclonación de las letras (Slant).

5.1. Interlineado

Para el interlineado se analizan dos clasificaciones posibles: Interlineado amplio (LINESPACE_SEPARATED) e Interlineado pequeño (LINESPACE_CROWDED). Para encontrar el interlineado se utilizó la técnica del perfil de proyección vertical suavizado (SPR) [7], esto permite separar las zonas de texto y espacios analizando los extremos del SPR, para las imágenes segmentadas es simple con este algoritmo determinar donde está el texto correspondiente a la línea y los espacios superior e inferior correspondientes al interlineado, así se puede determinar si una línea está muy unida a sus adyacentes considerando la proporción de píxeles de foreground (píxeles negros correspondientes a texto) que existen en las regiones identificadas como espacio.

Para analizar los SPR se separa la imagen en regiones verticales de tamaño 5 % del ancho de la imagen. Luego, se hace el SPR de cada uno, se determinan las regiones de texto y espacio [16] y se calcula la densidad de píxeles de foreground en los espacios. Finalmente, se promedian los resultados de densidades para cada sección vertical, para determinar si la línea está muy separada o muy pegada a sus adyacentes.

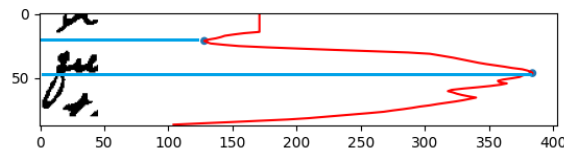


Figura 1: Segmento vertical de 5 % de imagen, en rojo el SPR y en azul los extremos del SPR, en el intermedio está la zona de texto y en los bordes la zona de interlineado.

5.2. Margen

Para determinar el margen se pueden considerar tanto los márgenes superior e inferior como los márgenes laterales [9]. Para nuestro Dataset, como las imágenes son líneas manuscritas, no tiene sentido analizar los

márgenes superior e inferior. Por tanto, solo se clasifica el margen lateral izquierdo en cuanto a su ancho como: Margen ancho (BIG_MARGIN) y margen estrecho (SMALL_MARGIN). Para esto se analizó el perfil de proyección (PR) horizontal de cada imagen tomando el rango del PR desde el elemento 0 hasta que el valor del PR sea mayor que una tolerancia especificada; es decir, mientras que la densidad de píxeles en el foreground para cada columna de la imagen sea menor que un valor relativo al largo de la imagen.

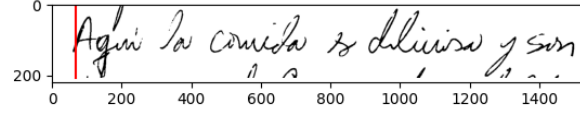


Figura 2: Desde el inicio hasta la línea roja la sección de margen.

5.3. Espaciado

El espaciado, son los espacios dejados entre caracteres y palabras en una misma línea. Para clasificar el espaciado se realizó a cada imagen un trabajo con filtros morfológicos; seleccionando un kernel horizontal para realizar una convolución con la imagen que permita unificar los trazos de todo lo que esté escrito unido de una misma palabra para de este modo separar las palabras enmarcándolas en recuadros. Así el espaciado se puede determinar como la distancia promedio entre los recuadros.

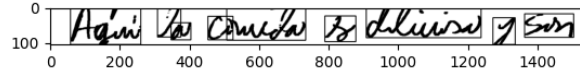


Figura 3: Recuadros que bordean al texto escrito sin espacios.

5.4. Inclinación de la Línea

Para determinar la inclinación de la letra en una línea se realizó el procedimiento descrito en [7], rotando la imagen desde -30° hasta 30° . Determinando para cuál de los valores de la rotación la cantidad de píxeles de foreground es máxima para la zona central de la imagen. Para los ángulos comprendidos entre $[-30; -5]$ se dice que la línea base es Ascendente (ASCENDING), entre $[-5; 5]$ es Nivelada (LEVELED) y entre $(5; 30]$ es Descendente (DESCENDING).

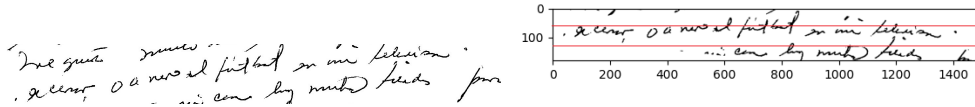


Figura 4: a) texto original b) Texto con una inclinación de -6° , la zona enmarcada en rojo es el área en que se busca maximizar la cantidad de pixeles de foreground.

5.5. Inclinación de la letra

Para la inclinación de la letra se tienen 5 posibles clasificaciones: Inclinado extremadamente a la izquierda (EXTREME_LEFT), ligeramente inclinado a la izquierda (MODERATE_LEFT), así como sus equivalentes hacia la derecha (EXTREME_RIGHT) y (MODERATE_RIGHT) respectivamente y la escritura sin inclinaciones (VERTICAL). Para extraer este feature se utiliza el perfil de proyección (PR) vertical rotando la imagen en ángulos entre -20° y 20° , buscando el ángulo que maximice la suma del PR normalizado, tal y como se plantea en [7].

6. Algoritmos Utilizados

Para la investigación sobre Grafología con aprendizaje de máquinas se utilizaron a modo comparativo los algoritmos más citados en el Estado del Arte de este tema. Para probar cada uno de ellos se dividió el dataset

en conjuntos: Entrenamiento/Test y Validación con tamaño 75 % y 25 % de las personas que conforman el dataset respectivamente. Primero se seleccionó el conjunto de personas para hacer la validación y se separaron las imágenes de sus escrituras, features extraídos y resultados del test BigFive. Luego a los restantes datos del dataset se les aplicó k-fold para dividir los datos de tests y entrenamiento.

6.1. KNN

Para el algoritmo de KNN se utilizaron los features extraídos de las imágenes y los resultados del big five para entrenar un modelo `KNeighborsClassifier` del módulo `sklearn` de `pyrhon`. Previo al entrenamiento se reescalaron los datos de la entrada utilizando `StandardScaler`. El valor de `k` fijado como cantidad de vecinos cercanos a fijar fue 7, con el cual se obtuvo los mejores resultados en la etapa de entrenamiento/test del algoritmo previo a la validación final. Cada característica se predijo por separado y se empleó una agrupación en intervalos de a 10.

6.2. CNN

En la Red Neuronal Convolutiva se utilizó directamente las matrices de píxeles de las imágenes de escrituras para entrenar el modelo. Primeramente se hizo un reescalado de la imagen para que todas tuviesen la misma resolución y luego se utilizaron en el entrenamiento de una CNN. Para la última capa de la CNN se tienen 5 neuronas, una por cada uno de los rasgos del BigFive a predecir.

Se probaron con varias ideas de entrenamiento, variando las dimensiones de las imágenes para determinar si la calidad afecta en mayor o menor medida la clasificación. No obstante teniendo en cuenta las características del dataset, en cuanto al volumen de imágenes el modelo con Redes Neuronales es propenso a overfit con pocas imágenes.

6.3. SVM

Para la máquina de soporte vectorial, como las clasificaciones según BigFive para cada uno de los rasgos puntúa con un valor de 0 a 120, y SVM separa por categoría, reescalamos los resultados de BigFive de 0 a 11, agrupando los valores del big five de 10 en 10. Como además los rasgos del Big Five son independientes, si con la grafología fuera posible determinarlos, su valor dependería solo de los rasgos de las letras y no uno de otro. Por tanto se creó una SVM por cada uno de los 5 rasgos de Big Five de modo que permita clasificar para cada persona los niveles de cada característica en que se encuentra. Para esto se entrenó el modelo `SVC` de `sklearn.svm` con los features extraídos de las fotos y con uno de los rasgos del BigFive en cada caso. Los datos fueron reescalados usando `StandardScaler` de `sklearn.preprocessing`. Como la clasificación es de más de 2 categorías se empleó una estrategia de clasificación de one-versus-one (ovo) y se probó el enfoque con los kernels: lineal, polinomial y rbf. De igual modo se hizo un análisis separando los rasgos en solo 2 categorías, Alto (1) y Bajo (0) y un análisis final intentando predecir los 5 rasgos en simultáneo.

6.4. K-Means

Se utilizó el modelo `KMeans` de `sklearn.clusters` con los datos procedentes de los features extraídos de las imágenes y los resultados del test BigFive, siguiendo dos enfoques para el análisis: a partir de los features determinar las características de la personalidad asociadas; y viceversa, a partir de los resultados del big five tratar de hallar los rasgos de la escritura que se repiten a través de los clústers. Para determinar el número de clústers `K` que mejor se ajusta a los datos se implementaron el método del codo para las medidas de inercia y distorsión, y el método de la silueta. Además se implementaron métodos para poder visualizar todos estos resultados en distintas dimensiones.

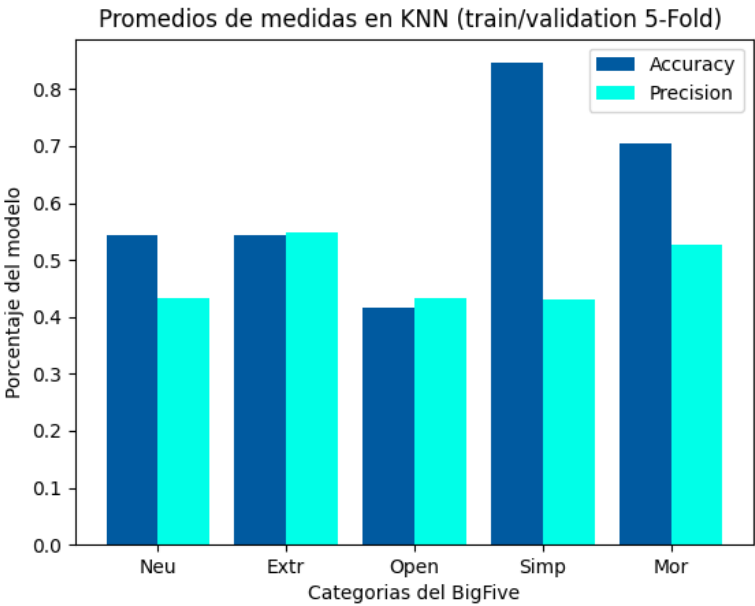
7. Resultados

7.1. KNN

En la evaluación del KNN se tuvieron en cuenta las métricas de `accuracy` y `precision`. Para cada uno de los rasgos por separado se obtuvo en la etapa de entrenamiento/test los siguientes valores promedio para

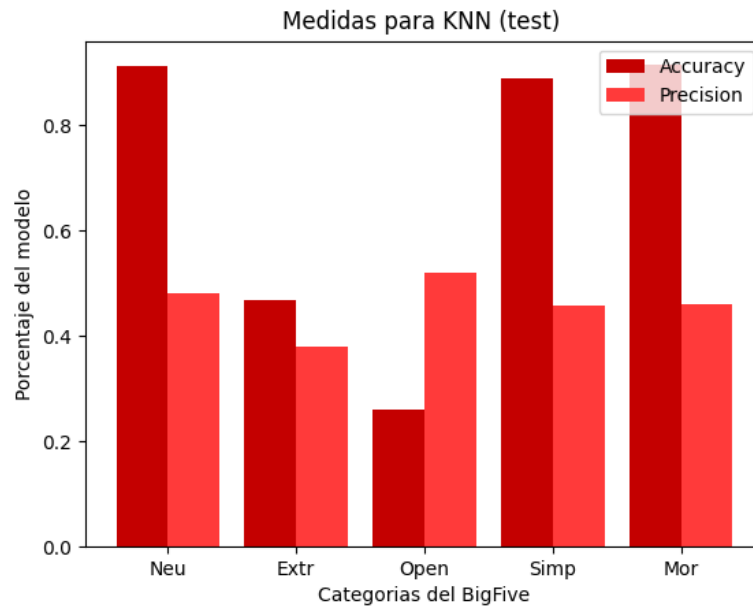
esas métricas luego de utilizar un 5-Fold:

Categoría	Accuracy promedio	Precisión promedio
Neuroticismo	0.5430	0.4334
Extroversión	0.5432	0.5473
Apertura a Experiencias	0.4155	0.4332
Simpatía	0.8458	0.4295
Moralidad	0.7042	0.5275



Posteriormente en la etapa de validación los resultados que se obtienen para **acurracy** y **precision** de cada rasgo son:

Categoría	Accuracy	Precisión
Neuroticismo	0.9108	0.4799
Extroversión	0.4682	0.3788
Apertura a Experiencias	0.2611	0.5211
Simpatía	0.8885	0.4574
Moralidad	0.9140	0.4614



7.2. CNN

Para la Red Neuronal Convolutiva se analizaron los rasgos del BigFive agrupando los valores de cada característica en bajo y alto nivel. Se tuvieron en cuenta para la evaluación del modelo las métricas: **mean absolute error** (MAE) y el **accuracy**. Los resultados para cada tamaño de entrada obtenidos son los siguientes:

Usando 3 capas de convolucion, reajuste de las imágenes a 32x32:

MAE para característica 0: 1.3867782231034904

MAE para característica 1: 1.304610701813095

MAE para característica 2: 0.9960057525342452

MAE para característica 3: 1.356362326391812

MAE para característica 4: 1.3710135627980433

Con un accuracy igual a 76.048 %.

Si se entrena con exactamente las mismas características pero con imágenes ajustadas a 64x64, para mayor calidad, obtenemos los siguientes resultados:

MAE para característica 0: 1.2410950697244811

MAE para característica 1: 1.0932612272971434

MAE para característica 2: 0.9566778752995634

MAE para característica 3: 1.0583139525519476

MAE para característica 4: 1.2694159266592442

Con un accuracy igual a 75.901 %.

Haciéndolo con dimensiones 128x128:

MAE para característica 0: 1.308632514485911

MAE para característica 1: 1.107600970286519

MAE para característica 2: 0.8781356299973996

MAE para característica 3: 1.008931401132167

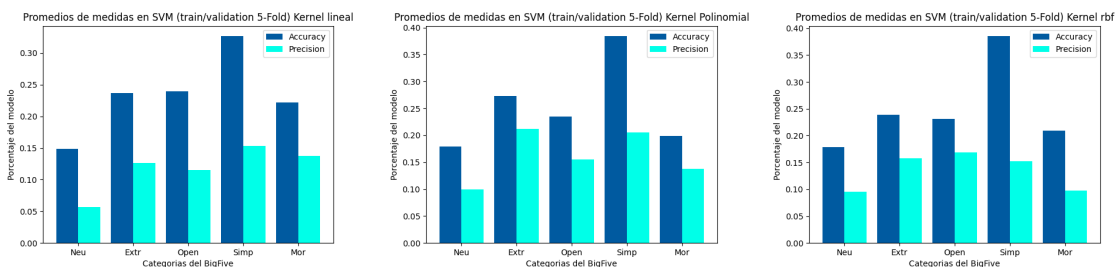
MAE para característica 4: 1.2417590243606276

Con un accuracy igual a 75.606 %.

Donde las características 1,2,3,4,5 son Neuroticismo, Extroversión, Apertura a Experiencias, Simpatía y Moralidad respectivamente.

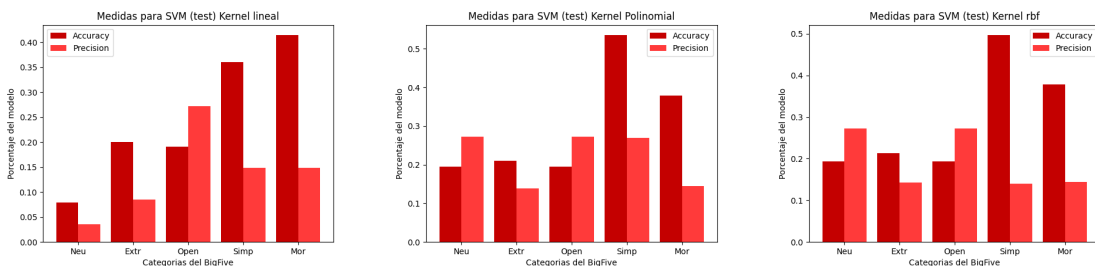
7.3. SVM

Para SVM con grupos de rango 10 probando con los tres tipos de kernel disponibles en `sklearn.SVM` en la etapa de entrenamiento/test, haciendo 5-Fold, se obtuvieron paracada razgo del Big Five los siguientes preomiedios para las medias de **accuracy** y **precision**.

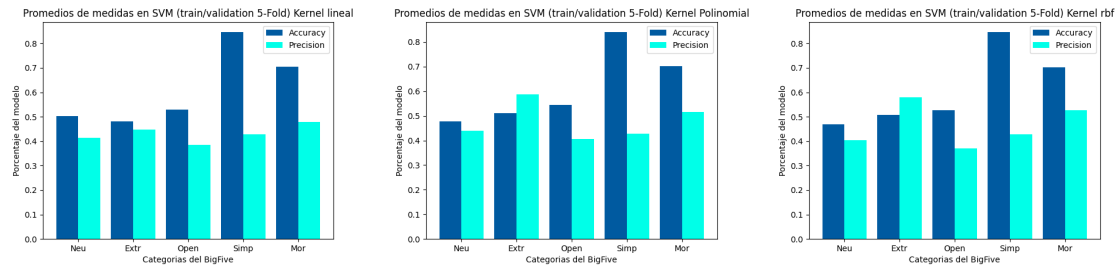


Todas por debajo del 50 %, para estos modelos en la etapa de validación se obtuvieron los siguientes valores de **accuracy** y las siguientes gráficas para cada kernel.

Categoría	Accur (Lineal)	Accur (Polinomial)	Accur (RBF)
Neuroticismo	0.0796	0.1401	0.1433
Extroversión	0.2006	0.2101	0.2133
Apertura a Exp	0.1910	0.1942	0.1942
Simpatía	0.3598	0.5350	0.4968
Moralidad	0.4140	0.3789	0.3789

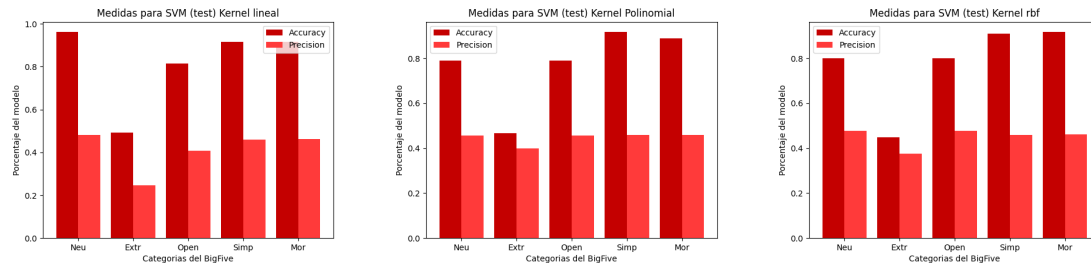


Además de este enfoque se trató también con una clasificación en dos categorías, Alta(> 80) y Baja(≤ 80) para cada rasgo del BigFive y se obtuvieron resultados que varían entre en 50 % y el 84 % de **accuracy** en la etapa de entrenamiento/test del modelo tal y como muestran las gráficas:

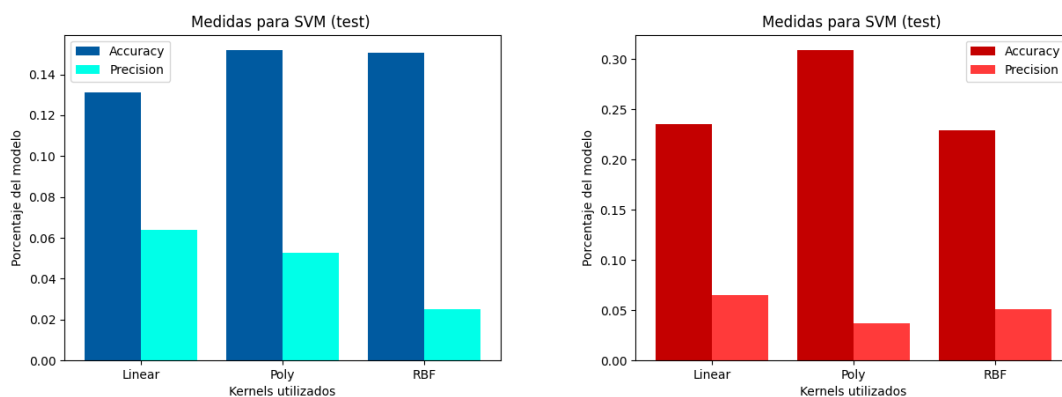


En la etapa de validación se obtuvieron los siguientes valores de **accuracy**:

Categoría	Accur (Lineal)	Accur (Polinomial)	Accur (RBF)
Neuroticismo	0.9617	0.9171	0.9171
Extroversión	0.4936	0.4681	0.4490
Apertura a Exp	0.8152	0.7898	0.7993
Simpatía	0.9171	0.9171	0.9108
Moralidad	0.9171	0.8885	0.9171



Por último se trató de predecir en simultáneo los 5 rasgos, para ello se tomó el vector binario resultante de clasificar cada categoría como alta o baja y se convirtió a decimal, con lo que se tienen 2^5 grupos posibles, como era de esperar debido a las dimensiones del dataset se obtuvieron bajos resultados para este enfoque debido a underfit, las gráficas de las etapas entrenamiento/test y validación se muestran a continuación, igual que los valores de **accuracy** en la etapa de validación final para el SVM con cada kernel.



Accuracy kernel lineal: 0.2365

Accuracy kernel polinomial: 0.3089

Accuracy kernel rbf: 0.2292

7.4. K-Means

El paso inicial para utilizar el método KMeans de sklearn es configurar el parámetro *random_state* el cual determina la generación de números aleatorios para la inicialización del centroide. De esta forma se puede asegurar que los resultados sean consistentes a través de múltiples ejecuciones. En dependencia de este valor podemos obtener distintos resultados a la hora de agrupar. Se hicieron pruebas para distintos valores de este parámetro, algunas de las cuales pueden ser encontradas en el directorio *KMeans_plots* del repositorio. Finalmente se decidió tomar *random_seed = 45*.

Luego, se aplican los métodos del codo y la silueta para un rango de clústers determinado, teniendo en cuenta ambos para poder escoger el número *k* óptimo de clústers. Además se grafican los datos después de ser agrupados para los mejores *k*. Se pueden encontrar para cada valor de *random_seed* todas las gráficas generadas por los métodos anteriores también en el directorio *KMeans_plots*. Finalmente, se toman los centroides de cada clúster, se agrupan las características correspondientes a cada dato según dichos clústers, y se toma su promedio como representativo del grupo para realizar inferencias.

7.4.1. Descubrir rasgos de caligrafía a partir de rasgos de personalidad

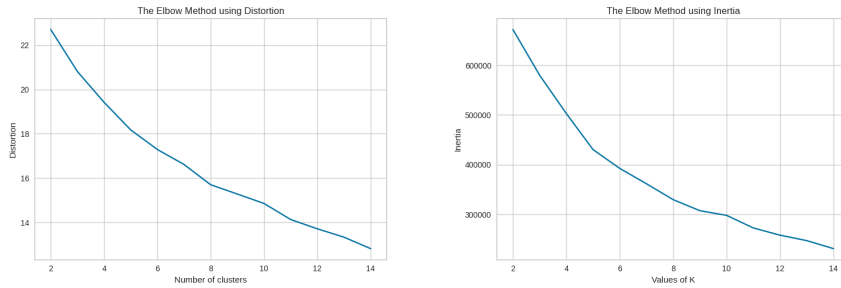


Figura 5: Método del codo para medidas distorsión (izquierda) e inercia (derecha) con número de clústers $k \in [2, 14]$.

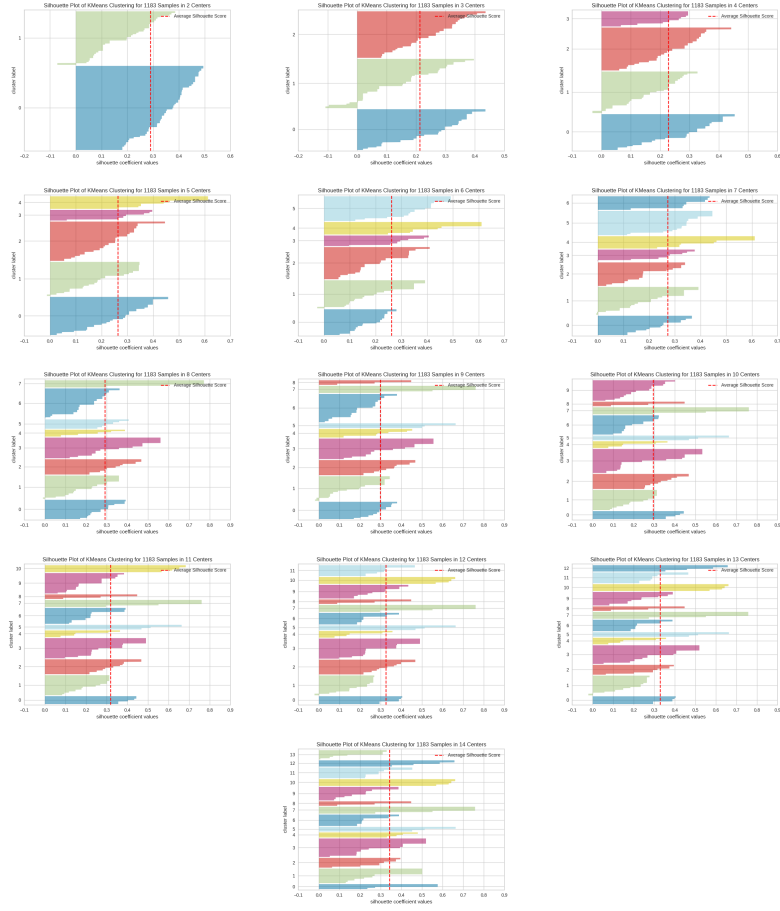


Figura 6: Visualización de la silueta para todo número de clúster $k \in [2, 14]$.

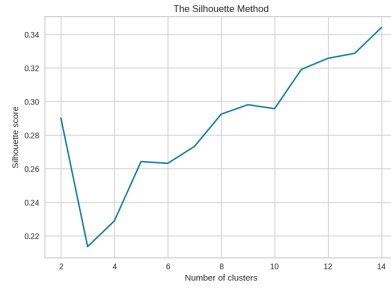


Figura 7: Método de la silueta con número de clústers $k \in [2, 14]$.

Teniendo en cuenta los resultados mostrados en las gráficas anteriores se determinó que el número óptimo de clústers es $k = 10$.

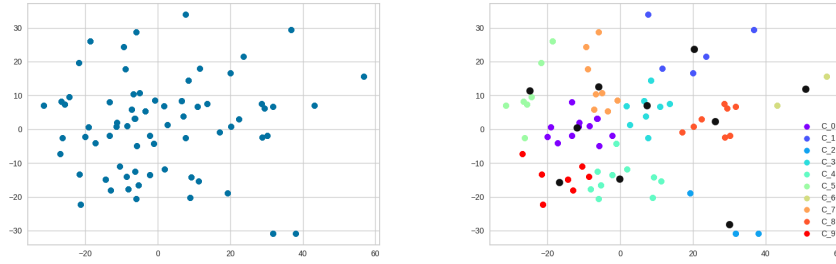


Figura 8: Datos originales (izquierda) y datos después de aplicarles clustering (derecha) para $k=10$ en dos dimensiones.



Figura 9: Datos originales (izquierda) y datos después de aplicarles clustering (derecha) para $k=10$ en tres dimensiones.

Dado al tamaño del dataset y la poca diversidad de los datos el modelo no es capaz de extraer grandes conclusiones. Nótese que el vector promedio de features por clúster es igual en todos excepto para el clúster número 7. De este podemos saber que las personas con baja Moralidad y todos los demás rasgos con valores altos, tienden a escribir con márgenes anchos.

Clúster	Promedio Big Five (centroides)	Promedio features
0	58. 70. 91. 95. 79.	0. 0. 1. 0. 0.
1	81. 74. 81. 78. 85.	0. 0. 1. 0. 0.
2	86. 91. 88. 97. 89.	0. 0. 1. 0. 0.
3	64. 94. 93. 97. 89.	0. 0. 1. 0. 0.
4	80. 56. 75. 92. 91.	0. 0. 1. 0. 0.
5	83. 71. 71. 88. 57.	0. 0. 1. 0. 0.
6	58. 85. 78. 97. 105.	0. 0. 1. 0. 0.
7	94. 92. 98. 86. 57.	0. 0. 1. 0. 1.
8	96. 49. 75. 81. 60.	0. 0. 1. 0. 0.
9	66. 76. 80. 94. 93.	0. 0. 1. 0. 0.

Figura 10: Tabla con los resultados al aplicar clustering para $k=10$. Se muestran las puntuaciones promedio de los datos referentes al big five y los features en cada clúster de la forma (Neuroticismo. Extroversión. Apertura a Experiencias. Simpatía. Moralidad.) y (Inclinación de la Letra. Inclinación de la Línea. Espaciado. Interlineado. Margen.) respectivamente.

7.4.2. Descubrir rasgos de personalidad a partir de features

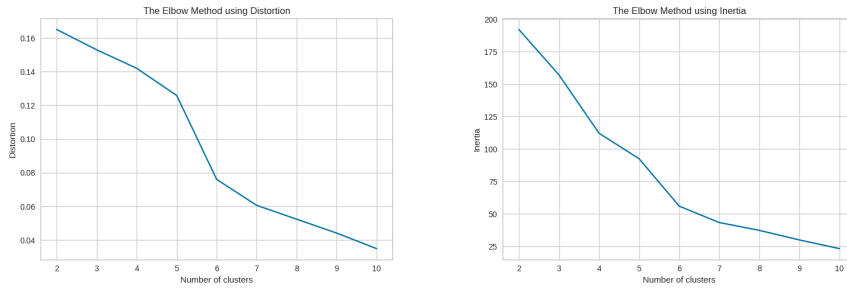


Figura 11: Método del codo para medidas distorsión (izquierda) e inercia (derecha) con número de clústers $k \in [2, 10]$.

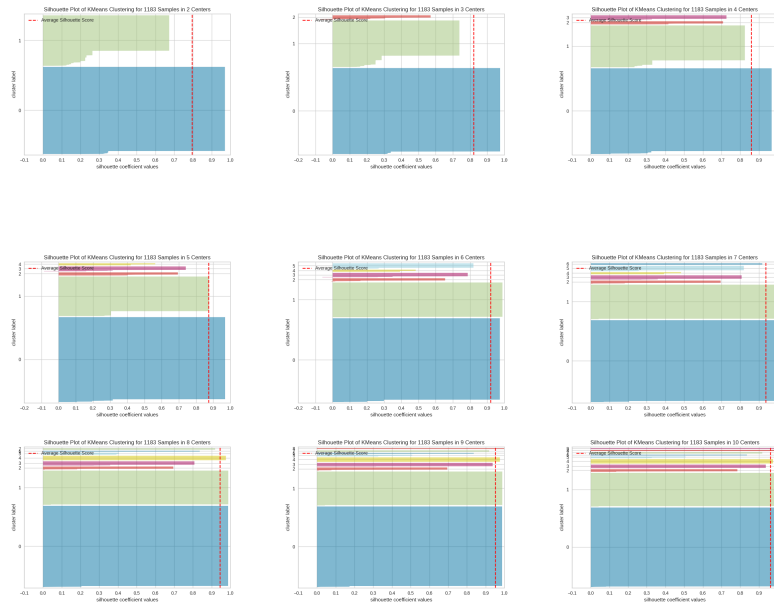


Figura 12: Visualización de la silueta para todo número de clúster $k \in [2, 10]$.

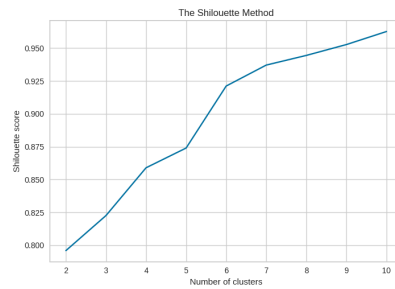


Figura 13: Método de la silueta con número de clústers $k \in [2, 10]$.

Teniendo en cuenta los resultados mostrados en las gráficas anteriores se determinó que el número óptimo de clústers es $k = 7$.

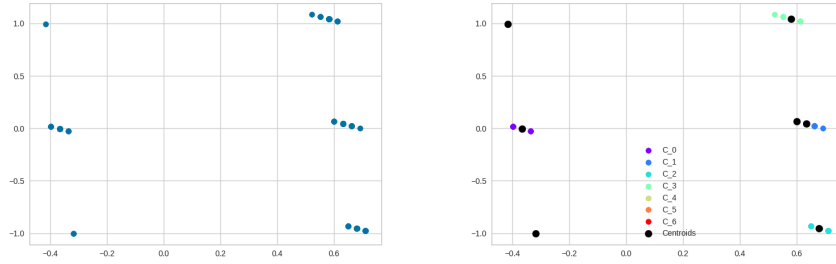


Figura 14: Datos originales (izquierda) y datos después de aplicarles clustering (derecha) para $k=7$ en dos dimensiones.

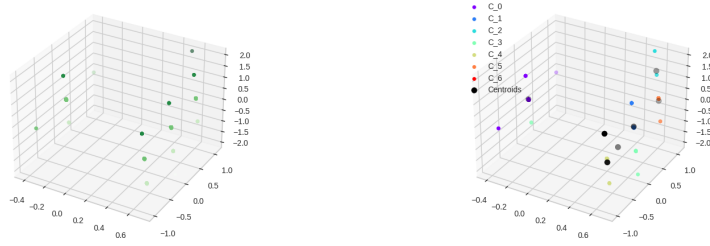


Figura 15: Datos originales (izquierda) y datos después de aplicarles clustering (derecha) para $k=7$ en tres dimensiones.

Al igual que en el enfoque anterior, se puede extraer poca información de los clústeres formados, viéndose poca diversidad en el dataset, por ejemplo aquí todas las personas escriben con espaciado estrecho. De forma general se puede inferir que las personas con interlineado pequeño tienden a un alto neuroticismo y una baja extroversión; y las personas que escriben con márgenes anchos, son usualmente poco neuróticas y muy extrovertidas, abiertas y simpáticas.

Clúster	Promedio features (centroides)	Promedio Big Five
0	0. 0. 1. 0. 0.	73. 79. 84. 92. 87.
1	0. 0. 1. 0. 1.	69. 83. 84. 92. 88.
2	1. 0. 1. 0. 1.	75. 81. 83. 92. 85.
3	0. 1. 1. 0. 1.	74. 81. 83. 87. 90.
4	1. 0. 1. 0. 1.	74. 84. 84. 91. 87.
5	0. -1. 1. 0. 1.	78. 81. 85. 87. 79.
6	0. 0. 1. 1. 0.	80. 74. 82. 96. 83.

Figura 16: Tabla con los resultados al aplicar clustering para $k=7$. Se muestran las puntuaciones promedio de los datos referentes a los features y el big five en cada clúster de la forma (Inclinación de la Letra. Inclinación de la Línea. Espaciado. Interlineado. Margen.) y (Neuroticismo. Extroversión. Apertura a Experiencias. Simpatía. Moralidad.) respectivamente.

8. Comparación de los algoritmos

Para comparar, entre los enfoques utilizados en la investigación, cuáles obtuvieron mejores resultados y por tanto permiten predecir mejor alguno de los rasgos del BigFive, o si con más de uno se obtienen resultados similares; se realiza para ello un análisis estadístico utilizando el 10-cross-fold-validation, es decir a cada algoritmo se le mide su rendimiento an diferentes datasets (10 creados por el 10-kfold) configurados con la misma semilla para mantener la uniformidad en el testeo. Con esto obtenemos para cada algoritmo, para

cada característica del BigFive las siguientes relaciones de **media** y **varianza** para la medida de accuracy, así como los gráficos de caja y bigotes para una comparación gráfica de los valores.

	Neuroticism		Extroversion		Exp. Openness		Sympathy		Morality	
	Media	SD	Media	SD	Media	SD	Media	SD	Media	SD
SVM Poly	0.5547	0.0311	0.5823	0.0329	0.5800	0.0268	0.8573	0.0374	0.7124	0.0526
	0.2819	0.0430	0.3107	0.0332	0.3464	0.0497	0.4223	0.0436	0.3199	0.0507
SVM Linear	0.5707	0.0313	0.5523	0.0244	0.5765	0.0314	0.8619	0.0332	0.7135	0.0480
	0.2658	0.0364	0.2946	0.0359	0.3556	0.0527	0.3866	0.0353	0.3176	0.0560
SVM rbf	0.5558	0.0310	0.5845	0.0314	0.5811	0.0282	0.8607	0.0329	0.7077	0.0522
	0.2750	0.0430	0.3084	0.0319	0.3544	0.0428	0.4269	0.0426	0.3187	0.0506
KNN	0.5466	0.0440	0.5638	0.0383	0.4246	0.0244	0.8504	0.0339	0.7077	0.0509
	0.0448	0.0096	0.2900	0.0415	0.3130	0.0300	0.4039	0.0328	0.2865	0.0586

Figura 17: Tabla de los algoritmos utilizados, para los valores de la media y desviación estandar del accuracy, para cada algoritmo se presentan las variantes de agrupar los datos en 2 categorías(fila superior) y la de agruparlos en 12 ctegorías(fila inferior) para cada uno de los 5 rasgos del BigFive

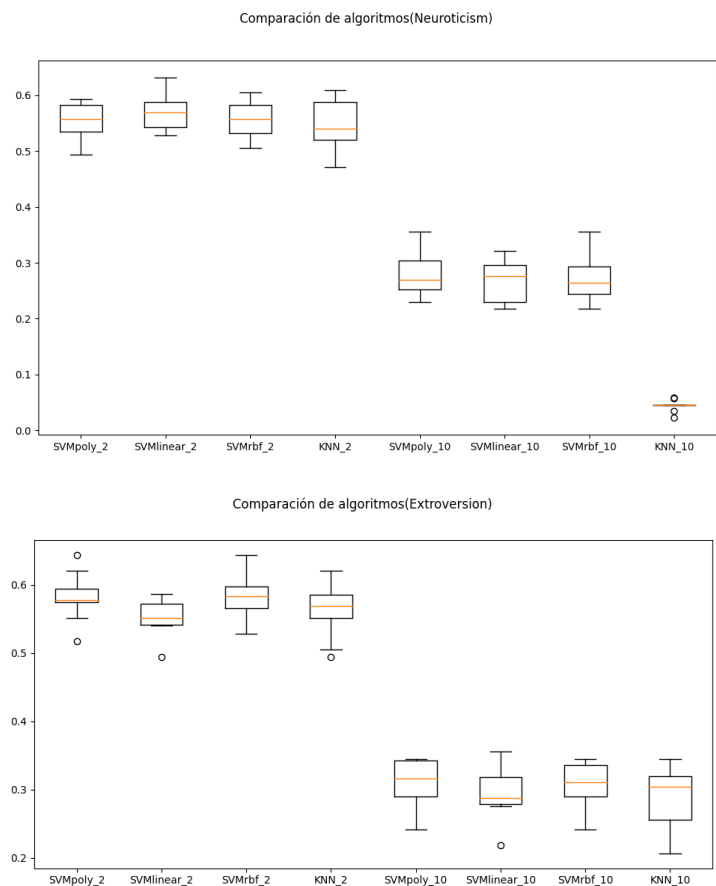


Figura 18: Gráficos de Caja y Bigotes de los resultados del 10-Kfold para los rasgos Neuroticismo y Extroversión

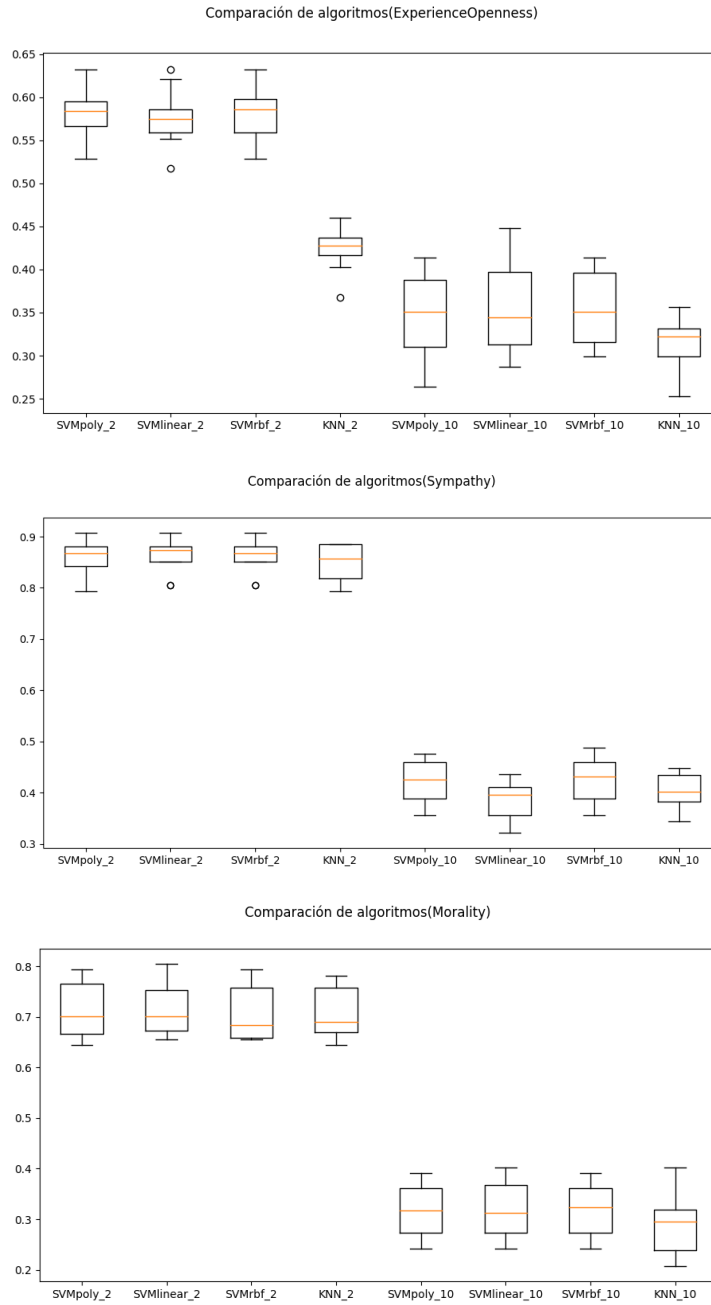


Figura 19: Gráficos de Caja y Bigotes de los resultados del 10-Kfold para los rasgos Apertura a experiencias, Simpatía y Moralidad

Con los valores de las medias, las varianzas y las gráficas se puede determinar que con gran diferencia los modelos para todas las características tienen un mejor desempeño cuando los datos están separados en dos categorías. No obstante, entre los algoritmos de KNN respecto a SVM, en algunas de las características del BigFive como la Apertura a experiencias, se puede comprobar que SVM obtiene un mejor resultado, mientras en las restantes queda la incertidumbre de si realmente es mejor utilizar algún kernel en específico para el SVM, o si es mejor usar el KNN. Precisamente para determinar si existen diferencias significativas entre estos algoritmos se les aplicó un test estadístico, el **t-test** que permite validar si las diferencias entre las métricas son o no estadísticamente significativas.

Esta prueba de hipótesis se encuentra en **Python** en el módulo **scipy.stats**, tiene las siguientes hipótesis nulas y alternativa:

H_0 : No existe diferencia significativa, las medias de las evaluaciones son iguales.

H_1 : Las medias de los valores son diferentes.

Al realizar estas pruebas para cada rasgo del BigFive, se obtuvieron los siguientes valores para el p -value.

	SVM-poly	SVM-linear	SVM-rbf	KNN
SVM-poly	-	0.066	0.588	0.284
SVM-linear	0.066	-	0.070	0.037
SVM-rbf	0.588	0.070	-	0.257
KNN	0.284	0.037	0.257	-

Figura 20: p-values del t-test para los algoritmos en el rasgo Neuroticismo

Tras la prueba de hipótesis, se tiene que para el SVM con kernel lineal, para todos los restantes algoritmos el p -value es significativamente pequeño con respecto a los restantes, aunque para un $\alpha = 0,05$ no se podría rechazar la hipótesis nula de su relación con los SVM de kernel polinomial y rbf, si se tomara un nivel de confianza ligeramente mayor si que se descartaría, además comparando la media y la varianza, se tiene que es el que mayor media obtiene y la varianza es relativamente igual a la de los restantes algoritmos, así que se puede afirmar que el algoritmo de SVM con kernel lineal es superior al resto para predecir el rasgo del Neuroticismo.

	SVM-poly	SVM-linear	SVM-rbf	KNN
SVM-poly	-	0.0005	0.3434	0.0107
SVM-linear	0.0005	-	0.0002	0.2579
SVM-rbf	0.3434	0.0002	-	0.008
KNN	0.0107	0.2579	0.008	-

Figura 21: p-values del t-test para los algoritmos en el rasgo Extroversión

Al realizar la prueba de hipótesis para el rasgo de Extroversión se genera entre los algoritmos dos grupos, por un lado están SVM de kernel rbf y polinomial y por otro en SVM de kernellineal y el KNN, entre los algoritmos de un mismo grupo con un nivel de confianza de un 95 % no es posible descartar la hipótesis nula, teniendose la opción de igualdad de medias, mientras que en los algoritmos intragrupo la hipótesis nula se descarta y hay diferencias significativas entre ellos. Al analizar los valores de ls medias y las varianzas se puede concluir la superioridad de los algoritmos SVM de kernel polinomial y rbf, siendo estos los mejores para predecir la extroversión.

	SVM-poly	SVM-linear	SVM-rbf	KNN
SVM-poly	-	0.557	0.726	0.000
SVM-linear	0.557	-	0.420	0.000
SVM-rbf	0.726	0.420	-	0.000
KNN	0.000	0.000	0.000	-

Figura 22: p-values del t-test para los algoritmos en el rasgo Apertura a experiencias

Para la apertura a experiencias, rasgo en el que se podía comprobar con mayor facilidad la diferencia entre los SVM y el KNN, el test ratificó dicha conclusión y además se obtiene que no se puede rechazar la hipótesis nula entre los tres tipos de kernels del SVM, siendo estos 3 algoritmos estadísticamente similares en la obtención de resultados para este rasgo del BigFive.

	SVM-poly	SVM-linear	SVM-rbf	KNN
SVM-poly	-	0.036	0.193	0.296
SVM-linear	0.036	-	0.343	0.062
SVM-rbf	0.193	0.343	-	0.121
KNN	0.296	0.062	0.121	-

Figura 23: p-values del t-test para los algoritmos en el rasgo Simpatía

Al analizar la tabla, se percibe que entre SVM de kernel lineal y el SVM de kernel polinomial, para un $\alpha = 0,05$ se puede descartar la hipótesis nula, de igual modo si fuéramos más restrictivos con el nivel de confianza, se podría incluso determinar que entre el SVM de kernel lineal y el KNN existen diferencias, no así entre los SVM linear y rbf, o entre el SVM plinomial y el KNN, en este rasgo en particular al analizar los gráficos vemos la presencia de outliers en las cajas bigote del SVM linear y d del rbf, que son los que influyen en que la varianza para los 4 algoritmos sea similar; de modo que se puede determinar que son ligeramente superiores estos dos al predecir la Simpatía.

	SVM-poly	SVM-linear	SVM-rbf	KNN
SVM-poly	-	0.783	0.167	0.308
SVM-linear	0.783	-	0.177	0.177
SVM-rbf	0.167	0.177	-	0.830
KNN	0.308	0.177	0.830	-

Figura 24: p-values del t-test para los algoritmos en el rasgo Moralidad

Todos los $p - values$ obtenidos son altos, por tanto para ninguna combinación de algoritmos se rechaza la hipótesis nula, quedndo que los 4 son relativamente igual de buenos en la predicción de la Moralidad.

En los casos en que la prueba de hipótesis **t-test** permitió rechazar la hipótesis nula, encontrándose diferencias entre las medias de las evaluaciones de los algoritmos, el procesos de decisión de cual de los algoritmos involucrados era el mejor se basó en analizar cuál tenía mejor media y menor varianza, o una varianza igual a las de los restantes algoritmos debido a que para cada uno de ellos se realizó un test de Normalidad de **Shapiro-Wilks**, en el que para cada uno se obtuvieron los siguientes $p - values$:

	Neuroticism	Extroversion	Exp. Openness	Sympathy	Morality
SVM Poly	0.4664	0.7526	0.9111	0.2827	0.1861
SVM Linear	0.6727	0.2556	0.7960	0.2051	0.3631
SVM rbf	0.8483	0.9970	0.5174	0.3412	0.0546
KNN	0.6947	0.6277	0.4317	0.0554	0.1234

Figura 25: Tablas de los p-value del test de normalidad para cada algoritmo en la predicción de cada uno de los rasgos del BigFive

Como para el test de **Shapiro-Wilks** se tiene

H_0 : Las muestras provienen de una distribución normal.

H_1 : No se cumple la normalidad en los datos.

Entonces con un nivel de confianza de un 95%, en ninguno de los algoritmos, para ninguna de las predicciones de rasgos del BigFive se puede rechazar la hipótesis nula, por lo que asumiendo el supuesto de Normalidad, quedan justificados las decisiones realizadas en la comparación de los algoritmos.

9. Conclusiones y trabajo futuro

En este trabajo se realizó un estudio profundo de aprendizaje de máquinas aplicado a Grafología, desde la consulta al Estado del Arte, Construcción y Taggeo de un Dataset, extracción de Features de las imágenes, entrenamiento de algoritmos y análisis de resultados. A modo general el flujo de trabajo queda representado en la siguiente imagen:

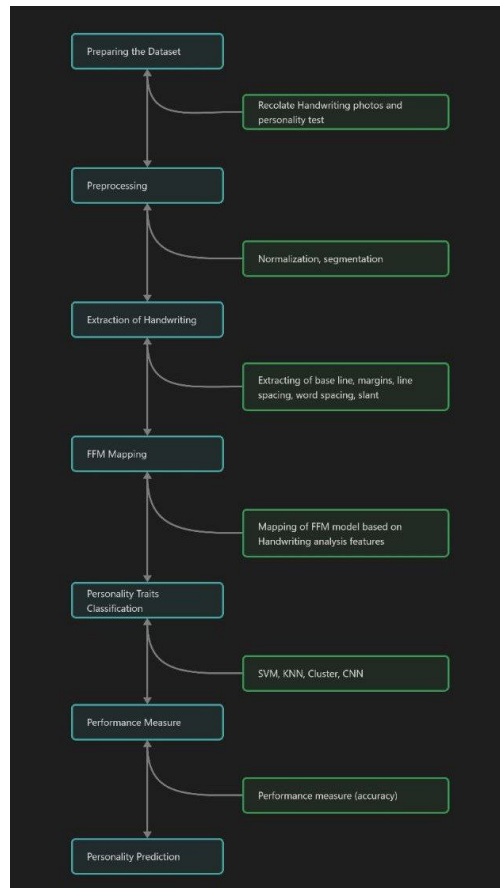


Figura 26: Flujo del trabajo del Proyecto.

Con los algoritmos empleados pudimos determinar que realmente la grafología no puede determinar de forma certera la psicología de una persona. Los modelos utilizados rondan el 50 % de efectividad en sus predicciones, lo cual dista de ser una clasificación fiable. De los modelos utilizados el SVM con clasificaciones en alto y bajo es el que mejores resultados ofreció, en su mayoría sobre el 80 % - 90 % de fiabilidad, pero igual de distante en identificar características como la extroversión.

Se propone como idea de investigación futura la construcción de un dataset de mayores dimensiones para probar estas y otras ideas de ML sobre grafología, que permitan que los modelos no hagan overfit por la escasez de datos respecto al amplio número de categorías para clasificar.

10. Repositorio del proyecto

<https://github.com/computer-science-crows/graphology-classification>

Referencias

- [1] Dataset 1. URL: <https://data.csafe.iastate.edu/DataPortal/#HandwritingDatabase>.
- [2] Dataset 2. URL: <https://fki.tic.heia-fr.ch/databases/iam-handwriting-database>.
- [3] Test de personalidad BigFive. URL: <https://bigfive-test.com>.
- [4] Yasemin Bay Ayzeren, Meryem Erbilek, and Erbuğ Çelebi. Emotional state prediction from online handwriting and signature biometrics. *IEEE Access*, 7:164759–164774, 2019.
- [5] HN Champa and KR AnandaKumar. Artificial neural network for human behavior prediction through handwriting analysis. *International Journal of Computer Applications*, 2(2):36–41, 2010.

- [6] Aditya Chitlangia and G. Malathi. Handwriting analysis based on histogram of oriented gradient for predicting personality traits using svm. *Procedia Computer Science*, 165:384–390, 2019. 2nd International Conference on Recent Trends in Advanced Computing ICRTAC -DISRUP - TIV INNOVATION , 2019 November 11-12, 2019. URL: <https://www.sciencedirect.com/science/article/pii/S1877050920300429>, doi:<https://doi.org/10.1016/j.procs.2020.01.034>.
- [7] Mihai Gavrilescu and Nicolae Vizireanu. Predicting the big five personality traits from handwriting. *EURASIP Journal on Image and Video Processing*, 2018:1–17, 2018.
- [8] Somayeh Hashemi, Behrouz Vaseghi, and Fatemeh Torgheh. Graphology for farsi handwriting using image processing techniques. *IOSR J. Electron. Commun. Eng.(IOSR-JECE)*, 10(3):01–07, 2015.
- [9] Prachi Joshi, Pranav Ghaskadbi, and Sarvesh Tendulkar. A machine learning approach to employability evaluation using handwriting analysis. In *Advanced Informatics for Computing Research: Second International Conference, ICAICR 2018, Shimla, India, July 14–15, 2018, Revised Selected Papers, Part I 2*, pages 253–263. Springer, 2019.
- [10] Rudi Kurniawan, Julian Supardi, Fatma Susilawati Mohamad, et al. A framework for determining the big five personality traits using machine learning classification through graphology. *Journal of Electrical and Computer Engineering*, 2023, 2023.
- [11] Nikita Lemos, Krish Shah, Rajas Rade, and Dharmil Shah. Personality prediction based on handwriting using machine learning. In *2018 International Conference on Computational Techniques, Electronics and Mechanical Systems (CTEMS)*, pages 110–113, 2018. doi:[10.1109/CTEMS.2018.8769221](https://doi.org/10.1109/CTEMS.2018.8769221).
- [12] Laurence Likforman-Sulem, Anna Esposito, Marcos Faundez-Zanuy, Stéphan Cléménçon, and Gennaro Cordasco. Emothaw: A novel database for emotional state recognition from handwriting and drawing. *IEEE Transactions on Human-Machine Systems*, 47(2):273–284, 2017.
- [13] Vaishali R Lokhande and Bharti W Gawali. Analysis of signature for the prediction of personality traits. In *2017 1st international conference on intelligent systems and information management (ICISIM)*, pages 44–49. IEEE, 2017.
- [14] I Maliki and MA Sidik. Personality prediction system based on signatures using machine learning. In *IOP Conference Series: Materials Science and Engineering*, volume 879, page 012068. IOP Publishing, 2020.
- [15] A Mishra. Forensic graphology: assessment of personality. *Forensic Res. Int. J*, 4(1):9–12, 2017.
- [16] Vassilis Papavassiliou, Themis Stafylakis, Vassilis Katsouros, and George Carayannis. Handwritten document image segmentation into text lines and words. *Pattern recognition*, 43(1):369–377, 2010.
- [17] Ajeet Ram Pathak, Abhishek Raut, Soham Pawar, Mansi Nangare, Harshdeep Singh Abbott, and Pawan Chandak. Personality analysis through handwriting recognition. *Journal of Discrete Mathematical Sciences and Cryptography*, 23(1):19–33, 2020.
- [18] Dian Pratiwi, Gatot Budi Santoso, and Fiqih Hana Saputri. Personality type assessment system by using enneagram-graphology techniques on digital handwriting. *International Journal of Computer Applications*, 147(11), 2016.
- [19] Ahmed Remaida, Aniss Moumen, Younes El Bouzekri El Idrissi, and Benyoussef Abdellaoui. Handwriting personality recognition with machine learning: a comparative study. In *2020 IEEE 2nd International Conference on Electronics, Control, Optimization and Computer Science (ICECOCS)*, pages 1–6. IEEE, 2020.
- [20] Pranoti Shete and Anita Thengade. E-graphologist for personality profile. *Asian Journal For Convergence In Technology (AJCT) ISSN-2350-1146*, 2019.
- [21] Sheetal Thomas, Mridula Goel, and Dipak Agrawal. A framework for analyzing financial behavior using machine learning classification of personality through handwriting analysis. *Journal of Behavioral and Experimental Finance*, 26:100315, 2020. URL: <https://www.sciencedirect.com/science/article/pii/S2214635019302291>, doi:<https://doi.org/10.1016/j.jbef.2020.100315>.

- [22] Nishigandha Vyawahare and Aparna Ashtaputre. Graphology- understanding and misunderstanding, 08 2022.