Universidad de Oviedo

Facultad de Ciencias

Predicción rápida y precisa de estructura electrónica con técnicas de aprendizaje automático

Claudio Sánchez Pérez de Amézaga.

Tutores:

Juan Luis Fernández Martínez.

Víctor Manuel García Suárez.

Grado de Matemáticas.

Julio 2020.

Índice

1. Predicción de materiales y el uso del *Machine Learning*.

El estudio de los materiales y la creación de nuevos compuestos es actualmente uno de los campos más fructíferos y con mayor relevancia. La búsqueda de materiales que se adapten a las diferentes necesidades tiene su origen hace miles de años con la fabricación de las primeras herramientas. Del uso de piedras se pasó a los primeros metales mejorando la tecnología de la época. Actualmente este proceso, ha pasado de una búsqueda de un material en la naturaleza, a su diseño. Y es que es más práctico diseñar el material que tenga unas propiedades concretas, a buscar en la naturaleza uno que se ajuste a nuestra necesidad.

El diseño de materiales no es tan sencillo actualmente, pero los resultados son revolucionarios. Las diferentes características que describen los materiales nos dan información sobre las propiedades que estos pueden tener. Por ejemplo el carbono puede formar tanto el grafito como el diamante. Dependiendo de cómo sea la estructura o red que forman los átomos, se obtienen unas propiedades u otras. El desarrollo del grafeno y los nanotubos de carbono son materiales revolucionarios que presentan propiedades de grandísimo interés, como dureza, flexibilidad o conductividad.

Con la introducción del *Machine Learning* en el desarrollo de materiales se pretende descubrir las relaciones entre los atributos fácilmente medibles de un sistema y sus propiedades. Los elementos de la tabla periódica y su comportamiento poseen atributos que pueden ayudar a predecir propiedades buscadas. Algunos de estos atributos son sencillos de medir, como el número de protones, el número de electrones en las capas más externas del átomo, o el tamaño promedio del átomo. Otros son más complicados, como por ejemplo la electronegtividad. Sin embargo, las técnicas de *Machine Learning* también permiten predecir estas variables difíciles de determinar.

Las combinaciones de estos atributos generan un gran número de posibilidades, así que no es de extrañar que materiales que involucran un número reducido de elementos de la tabla periódica, pueda formar una enorme cantidad de compuestos diferentes, con propieda-

des que aún no se han descubierto. En esta situación, surge la necesidad de contar con un método adecuado para predecir de manera rápida y precisa compuestos aún por sintetizar. Actualmente los métodos de descubrimiento cuentan con procesos experimentales costosos y largos, o cálculos intensivos en casos particulares. Estas técnicas son incapaces de encontrar las reglas semi empíricas que rigen el comportamiento de los materiales.

1.1. Perfil de los materiales.

Todas las propiedades o atributos de un material componen su perfil o *finger print*. Al igual que una huella dactilar pertenece a una única persona, este perfil del material, identifica cada material por separado. A partir de esta descripción, se pretende realizar un mapeo entre su perfil, y las propiedades que posee un material. Estas propiedades son conocidas debido a razones históricas, como su dureza o conductividad, o también, pueden ser generadas intencionalmente.

La naturaleza de estos atributos corresponde a descriptores químicos y físicos. Algunos de ellos se conocen desde hace mucho tiempo, mientras que otros son completamente nuevos. No se puede descartar cómo influye un aspecto concreto en el material final que compone dicho material. La identificación de estos atributos es de vital importancia. Conocerlos puede suponer un problema en si mismo. La gran variedad de descriptores de un elemento hace que se requiera una selección previa. Aunque trabajar con un gran número de descriptores puede suponer una ventaja, requiere una mayor comprensión del elemento, que no siempre es fácil. En este trabajo diferenciaremos en dos clases de atributos.

El primer conjunto de propiedades que conforman el perfil del elemento son características que poseen los elementos de la tabla periódica en su naturaleza. Entre estos se incluyen el número de protones, o el radio de sus orbitales. Son características por tanto universales y que ya han sido estudiadas y tabuladas. El segundo conjunto de atributos corresponden a características estructurales. La estructura que forma un sólido no es arbitraria y tiene una gran influencia sobre las propiedades del material. La principal diferencia entre el grafito y

el diamante es precisamente cómo se colocan los átomos de carbono. Las estructuras principales que se encuentran en la naturaleza se exponen más adelante y son lo suficientemente influyentes como para poner especial atención. De esta forma se obtienen atributos que no solo dependen de cada elemento, si no que dependen de la estructura en la que se encuentren. La obtención de estos valores pasa por tanto por la densidad de estados que será introducida más adelante y juega un papel fundamental en este trabajo. Estos valores se han calculado y analizado en esta investigación, obteniéndolos de manera directa.

1.2. Objetivo de la investigación.

El objetivo que tiene esta investigación es conocer y aplicar técnicas de *Machine Learning*. Los análisis a través de herramientas matemáticas tiene cada vez más fuerza, respaldados por su eficiencia y eficacia. Se pretende por tanto aplicar esta rama a un caso concreto de gran importancia como es el desarrollo de nuevos materiales. Se ha optado por una procedimiento en el que el aprendizaje de nuevas herramientas va de la mano con su aplicación inmediata y un análisis inmediato en un problema actual.

Esta investigación sigue el proceso desde la elaboración del perfil de los elementos hasta la realización del mapeo entre el perfil y la propiedades. De esta forma la predicción de nuevas propiedades dentro de materiales de la misma clase tendrá un coste computacional infinitesimal comparado con los costes experimentales tradicionales. Este desarrollo está motivado por la idea de que materiales similares, tendrán propiedades similares.

El problema central en este problema de aprendizaje pasa por establecer un mapeo entre los atributos fáciles de obtener que componen el perfil de los materiales y sus propiedades.

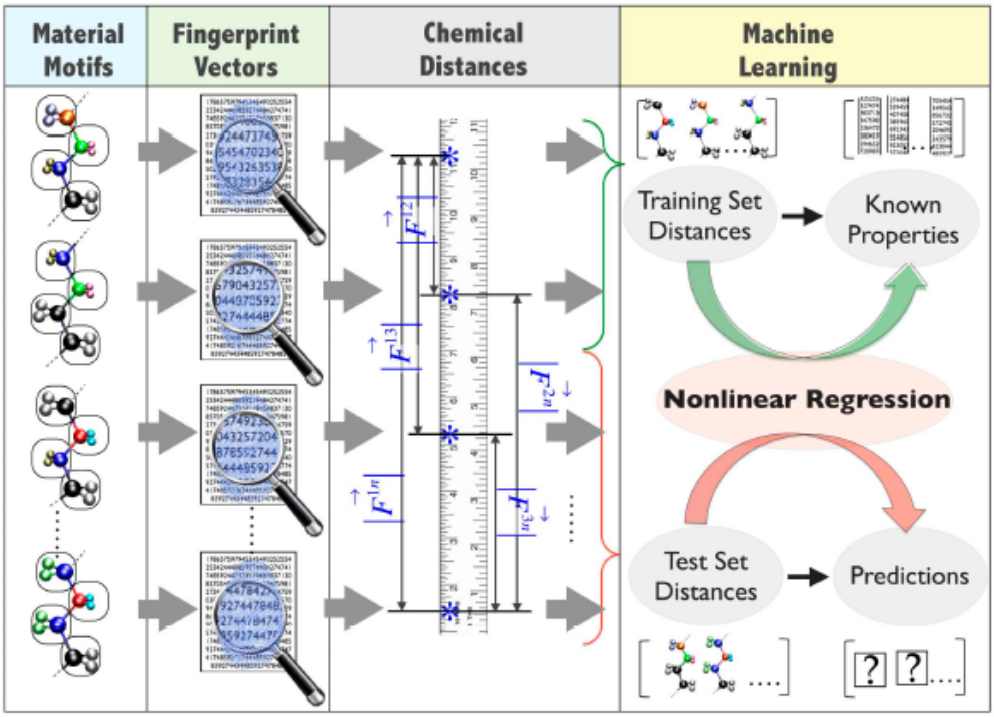

Figura 1: Método de *Machine Learning* [1].

La figura (1) representa el proceso de *Machine Learning* para la predicción de propiedades de nuevos materiales. El primer paso es el tratamiento de los atributos, que se convierten en vectores numéricos que identifican el material. A continuación se realiza una clasificación de los elementos para finalmente predecir sus propiedades. La clasificación de los elementos nos permite establecer qué propiedades puede tener un material, sabiendo que sus componentes pertenecen a una clase de la que conocemos sus propiedades.

La clasificación de los elementos es de vital importancia y no se hace de manera arbitraria. Por ejemplo, en la tabla periódica, encontramos juntos a los metales o los gases nobles, quienes comparten características. La clasificación es por tanto un problema en si mismo. Puede realizarse de una manera subjetiva o por el contrario seguir el mismo criterio en todos los casos. Este trabajo también incluye un análisis sobre la clasificación, que es utilizado a lo largo de toda la investigación.

Esta investigación parte de tres problemas que se han introducido anteriormente:

- La identificación de atributos y la elaboración del perfil de cada elemento. A partir de la base de datos con la que trabajamos, obtenemos los atributos naturales directamente y los estructurales gracias a análisis de la base de datos. Es problema incluye la predicción de la electronegatividad como ejemplo de la obtención de un atributo a partir del resto.

- La predicción de la densidad de estados de los elementos. La densidad de estados juega un papel fundamental en el estudio de los sólidos en el las propiedades físicas de la materia. Su obtención no es inmediata y el uso del *Machine Learning* ayudará a su obtención computacional reduciendo costes experimentales.

- Por último, el uso de algoritmos de clasificación en este trabajo, motivó el desarrollo de un algoritmo de optimización de grupos que ha sido utilizado en el transcurso de la investigación, dando así una mayor solidez a los análisis.

2. Introducción Teórica.

La Física del Estado Sólido pertenece a la Física de la Materia Condensada, y estudia las propiedades de los materiales, desde sus capacidades térmicas, conducción eléctrica o la disposición de los átomos que dotan a los materiales de diferentes propiedades. A continuación, se presentan unas definiciones previas de conceptos básicos en este ámbito, que nos ayudarán a entender mejor el problema y a enfocar el análisis.

2.1. Estructuras.

La disposición de los átomos en un sólido es esencial a la hora de determinar sus propiedades, por eso la estructura que forman es de vital importancia. El concepto fundamental estructural en la Física de Materiales es la Red de Bravais.

1. Red de Bravais (RB). Una red de Bravais es una distribución periódica de puntos o nodos, en la que todos los puntos, o nodos, son equivalentes. Es decir, todos los puntos están rodeados por el mismo entorno. De esta forma podemos encontrar simetrías como, traslaciones o transformaciones puntuales, como rotaciones, planos de simetría e inversión espacial.

2. Redes de Bravais en tres dimensiones. En el desarrollo de esta investigación, nos centraremos en cuatro redes de Bravais tridimensionales, estas son la SC, BCC, FCC y la HCP.

    - *Simple Cubic*, SC. Se trata de una RB compuesta por un cubo en el que cada vértice está ocupado por un nodo.

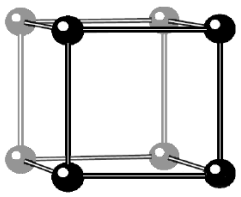

Figura 2: Estructura SC.

- *Body Centered Cubic*, BCC. Tiene la misma estructura que la SC, añadiendo un nodo en el centro del cubo.

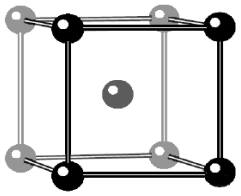

Figura 3: Estructura BCC.

- *Face Centered Cubic*, FCC. Es una estructura SC, en la que se añade un nodo en el centro de cada una de las seis caras del cubo.

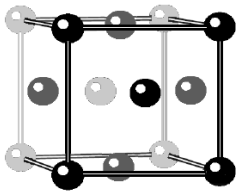

Figura 4: Estructura FCC.

- *Hexagonal Closed Packed*, HCP. Es una estructura más comlpleja, que puede ser descrita a través de capas. Si consideramos capas paralelas al plano XY de la figura (5), donde los planos contienen exágonos regulares con nodos en los vértices y el centro de dicho exágono, las capas se van superponiendo de la forma AC AC indefinidamente siendo A igual a C desplazada de forma que la proyección en el plano XY del nodo de la capa superior coincide en el centro de los triángulos equiláteros formados por los exágonos.

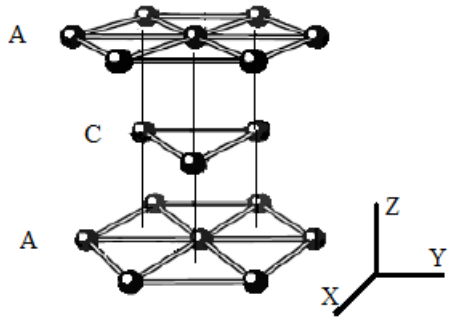

Figura 5: Estructura HCP.

La estructura FCC guarda una relación con la HCP, ya que ambas pueden definirse como una superposición de capas de la forma AC AC en la HCP y ACB ACB en la FCC. La capa B es la que hace diferentes ambas capas, pero puede ser útil a la hora de analizar los resultados pudiendo obtener datos similares.

3. Celda unidad. Región del espacio que, repetida mediante traslación, genera todo el espacio de la red de Bravais. Se trata de la fracción más pequeña de la red. La distancia constante entre las celdas unidades se conoce como parámetro de red, o constante de red. Las estructuras tridimensionales poseen tres parámetros de red, pero en el caso de la SC, la FCC, la BCC, los tres parámetros coinciden, ya que se trata del valor de la arista del cubo exterior que las compone. En el caso de la HCP, la constante de red es la distancia entre dos átomos consecutivos contenidos en el mismo plano. Se trata del valor de la arista del hexágono de la capa A de la figura (5).

4. Primeros vecinos. Se denominan nodos primeros vecinos de un nodo fijo, a aquellos que están más cerca de él, es decir, es la distancia más pequeña que separa dos nodos. Los siguientes nodos más cercanos se denominan segundos vecinos y así sucesivamente.

2.2. átomos y electrones.

Los átomos y los electrones son los ladrillos que componen la estructura final, por eso las propiedades de estos juegan un papel decisivo a la hora de buscar nuevas propiedades.

2.2.1. Los cuatro números cuánticos.

La disposición de los electrones en un átomo no es aleatoria, y sigue unas reglas muy estrictas que determinan su posición y su energía. A partir de la ecuación de Schrödinger, nos da información de la energía de un estado a través de su Hamiltoniano[1]. Y es en los posibles valores de la energía donde surgen los números cuánticos.

- Número cuántico principal n. El número cuántico principal nos dice el nivel de energía de un orbital[2]. Puede tomar valores desde 1, el valor menos energético, hasta infinito.
- Número cuántico azimutal u orbital l, determina el momento angular de un electrón en el orbital. Puede tomar valores desde 0 hasta n-1. Habitualmente se identifica cada número de l con una letra, siendo s para l=0, d para l=2, d para l=2, f para l=3 y sigue con el orden alfabético a partir de l=4.
- Número cuántico magnético m_l, indica la orientación del orbital en el espacio, y puede tomar valores desde -l...0..+l.
- Número cuántico de espín m_s, indica el sentido de rotación del electrón en el orbital. Solo hay dos posibilidades de giro, siendo estas $\frac{-1}{2}$ y $\frac{+1}{2}$.

Los cuatro números cuánticos especifican completamente la posición de un electrón en un átomo y siguen el principio de Exclusión de Pauli, que en el caso de los electrones (fermiones[3]), establece que no pueden haber dos electrones con los cuatro números cuánticos idénticos en el mismo átomo. Esto provoca que la posición de los electrones siga el siguiente esquema de llenado:

La figura (6) establece cómo van ocupando las posiciones los electrones, empezando por los niveles menos energéticos, hasta los más energéticos. Los electrones que tienen el mismo número principal n, se dicen que están en la misma capa. Cada capa puede contener $2n^2$ electrones, y si alcanza el número máximo, se dice que la capa esta llena. Los electrones que

[1]Función que describe la energía de un sistema físico.

[2]Región del espacio determinada por una solución de la ecuación de Schrödinger espacial e independiente del tiempo para un electrón sometido a un potencial Culombiano.

[3]Tipo de partícula elemental del modelo estándar que constituyen la materia.

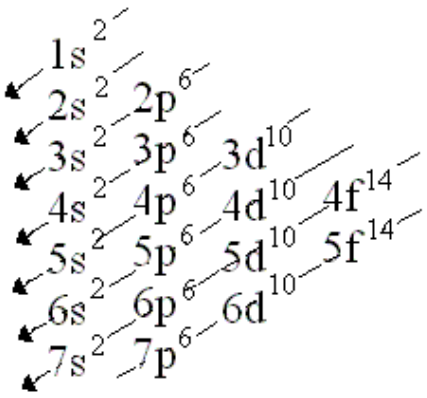

Figura 6: Esquema de la posición de los electrones en un átomo.

comparten número cuántico n y l se dicen que están en la misma subcapa. Podemos ver en la figura (6) que las subcapas no se van llenando ordenadamente a medida que aumenta n, es que hay subcapas con un n mayor, menos energéticas que otras con n menor. Por ejemplo un electrón ocuparía antes una posición 4s que una 3d. Cabría esperar que cada electrón en el átomo tuviese una energía concreta, determinada por sus números cuánticos. Sin embargo, la energía está degenerada para los números cuánticos m_l y m_s. Esto hace que los átomos con los mismos n y l tengan la misma energía. La degeneración de la energía para n y l fijos es $2(2l+1)$, que son todos los posibles valores de m_l y m_s.

2.3. Densidad de estados.

La densidad de estados, *Density of states* (DOS), es un concepto clave en la física de la materia condensada y juega un papel importante en esta investigación. La densidad de estados representa el número de estados por unidad de energía en un sistema físico.

A continuación, a modo de ejemplo, se presenta la densidad de estados más sencilla, la del Hidrógeno, en la que sólo interviene la capa n=1, l=0, es decir, la capa 1s.

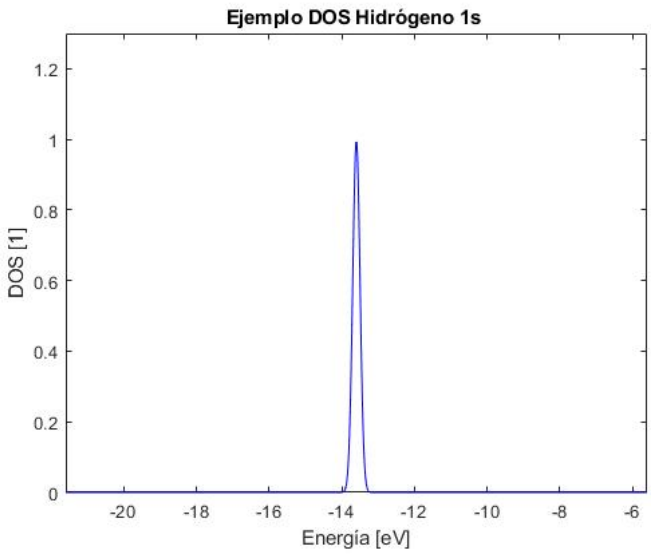

Figura 7: DOS para el átomo de Hidrógeno representado en unidades arbitrarias de la energía.

En el caso de un átomo que involucrase la segunda capa, es decir $n = 2$, se pasa a considerar 2 subcapas, los orbitales 2s y 2p. Ahora surge un nuevo pico, con un máximo hasta 4 veces superior, ya que en el nivel uno puede haber como mucho 2 electrones, mientras que en el nivel 2 puede haber un máximo de 8 (2 del orbital s y 6 del p).

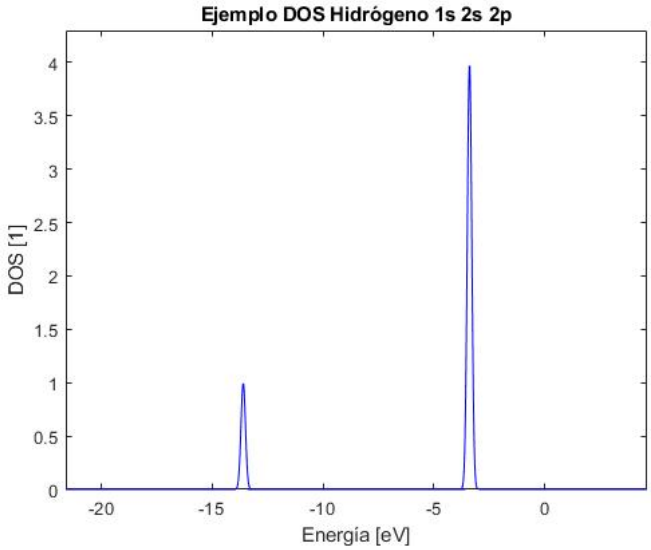

Figura 8: DOS del Hidrógeno considerando un llenado en el nivel energético 1 y 2.

Para definir la densidad de estados, consideramos un sistema de electrones libres. Esto

quiere decir que los electrones no interactúan entre ellos, ni con impurezas ni otros factores. Los electrones se encuentran con un potencial químico μ [4]. Entonces la probabilidad de que un estado de energía E esté ocupado, está dada por la expresión:

$$n_F(\beta(E-\mu)) = \frac{1}{e^{\beta(E-\mu)}+1} \tag{1}$$

donde $\beta=\frac{1}{K_B T}$, siendo K_B es la constante de Boltzman (K_B=1.38064852 x 10^{-23} J/K).

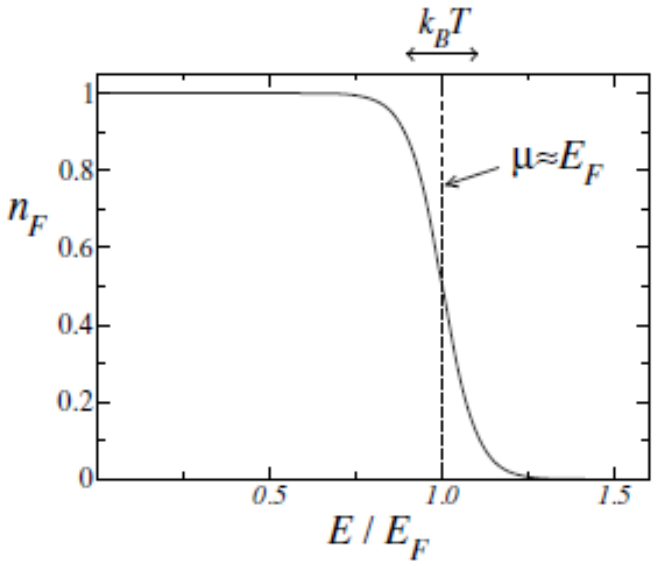

Figura 9: Representación de la ecuación (1) de la distribución de Fermi cuando $k_BT \ll E_F$. E_F es la energía de Fermi introducida a continuación [3].

Consideramos que los electrones se encuentran en una caja de lado L, y volumen V=L^3, que cuenta con condiciones periódicas en las fronteras. Las funciones de onda correspondientes a los estados tienen la forma $e^{i\mathbf{k}\cdot\mathbf{r}}$. Dadas las condiciones de contorno $\mathbf{k}$ toma valor $(2\pi/\mathrm{L})(n_1, n_2, n_3)$, con n_i un valor entero. Las energías de estos estados es:

[4]El potencial químico en un sólido representa una función de localización espacial, en la que los electrones o moléculas tienden a desplazarse desde regiones con alto potencial químico a regiones de bajo potencial. También puede interpretarse como el nivel que separa los estados llenos, de los vacíos. Su sus unidades son de energía por partícula. Si U es la energía interna y N el numero de partículas, entonces $\mu=\partial \mathrm{U}/\partial \mathrm{N}$ a volumen y entropía constantes.

$$\epsilon(\mathbf{k}) = \frac{\hbar^2|\mathbf{k}|^2}{2m} \tag{2}$$

siendo m la masa del electrón. El número total de electrones en el sistema viene dado por:

$$N = 2\sum_{\mathbf{k}} n_F(\beta(\epsilon(\mathbf{k}) - \mu)) = 2\frac{V}{(2\pi)^3}\int d\mathbf{k}\ n_F(\beta(\epsilon(\mathbf{k}) - \mu)). \tag{3}$$

El factor 2 inicial considera las dos posiciones del spin para cada vector de onda $\mathbf{k}$. En este punto podemos definir la energía de Fermi E_F como, el potencial químico a temperatura T=0 K.

A partir de la energía de Fermi unidimensional, se define la temperatura de Fermi T_F=E_F/k_B, y el vector de onda de Fermi, k_F cumpliendo:

$$E_F = \frac{\hbar^2 k_F^2}{2m}, \tag{4}$$

calculamos ahora la energía de Fermi en el caso tridimensional, en un sistema de N electrones. La función de Fermi (1) pasa a ser una función escalón Θ, donde Θ(x)=1 para x $\geq$ 0 y θ(x)=0 para x < 0. De esta forma la ecuación (3) pasa a:

$$N = 2\frac{V}{(2\pi)^3}\int d\mathbf{k}\ \Theta(E_F - \epsilon(\mathbf{k})) = 2\frac{V}{(2\pi)^3}\int^{|\mathbf{k}|<k_F} d\mathbf{k}. \tag{5}$$

La integral final es una integral sobre una bola de radio k_F, así que el resultado es el volumen de la bola:

$$N = 2\frac{V}{(2\pi)^3}\left(\frac{4}{3}\pi k_F^3\right). \tag{6}$$

Esta expresión nos dice que a temperatura T=0, los electrones llenan una bola en el espacio k de radio k_F. Sabiendo que la densidad es $n = N/V$, obtenemos:

$$k_F = (3\pi^2 n)^{\frac{1}{3}}. \tag{7}$$

De lo que deducimos:

$$E_F = \frac{\hbar^2(3\pi^2 n)^{\frac{2}{3}}}{2m} \tag{8}$$

Llegados a este punto, la energía total de un sistema de electrones viene dado por:

$$E_{total} = 2\frac{V}{(2\pi)^3}\int d\mathbf{k}\ \epsilon(\mathbf{k})\ n_F(\beta(\epsilon(k)-\mu) = 2\frac{V}{(2\pi)^3}\int_0^\infty 4\pi k^2\ \epsilon(\mathbf{k})\ n_F(\beta(\epsilon(\mathbf{k})-\mu). \quad (9)$$

En esta ecuación hemos cambiado a coordenadas esféricas para pasar a un problema unidimensional con un factor $4\pi k^2$. Ahora podemos reescribir la ecuación(9) reemplazando k por la energía ϵ:

$$k = \sqrt{\frac{2}{\hbar^2}} \qquad dk = \sqrt{\frac{m}{2\epsilon\hbar^2}}d\epsilon \quad (10)$$

de donde obtenemos:

$$E_{total} = V\int_0^\infty d\epsilon\ \epsilon\ g(\epsilon)\ n_F(\beta(\epsilon-\mu)) \quad (11)$$

$$N = V\int_0^\infty d\epsilon\ \ g(\epsilon)\ n_F(\beta(\epsilon-\mu)). \quad (12)$$

A $g(\epsilon)\ d\epsilon$ se denomina densidad de estados y representa el número total de estados con energía entre ϵ y $\epsilon + d\epsilon$.

$$g(\epsilon)d\epsilon = \frac{2}{(2\pi)^3}4\pi k^2 dk = \frac{2}{(2\pi)^3}4\pi\left(\frac{2\epsilon m}{\hbar^2}\right)\sqrt{\frac{m}{2\epsilon\hbar^2}}d\epsilon =$$

$$= \frac{(2m)^{\frac{3}{2}}}{2\pi^2\hbar^3}\epsilon^{\frac{1}{2}}d\epsilon, \quad (13)$$

introduciendo la energía de Fermi, E_F, podemos expresar la densidad de estados como:

$$g(\epsilon) = \frac{3n}{2E_F}\left(\frac{\epsilon}{E_F}\right)^{\frac{1}{2}}. \quad (14)$$

Las unidades de la densidad de estados es de densidad dividido por energía. Como la densidad de estados es una curva continua, si realizamos la integral con límite inferior -∞ y límite superior la energía de Fermi, E_F, obtenemos el número de partículas totales del sistema [3].

2.4. Base de Datos

La base de datos con la que iniciamos esta investigación se divide en cuatro grupos principales. Cada grupo se corresponde con cada una de las cuatro estructuras tridimensionales ya mencionadas, la SC, la FCC, la BCC y la HCP. Además, para cada estructura se utilizan

los datos de las DOS calculada con una constante de red fijada, es decir, todos los elementos forman la misma estructura con la misma constante de red, y con la constante de red más estable de cada elemento en la naturaleza, por lo que cada elemento formará la misma estructura pero con una constante de red independiente. En cada estructura analizamos 30 elementos. Éstos son los primeros 30 metales de transición, tal y como se indica en la figura (10):

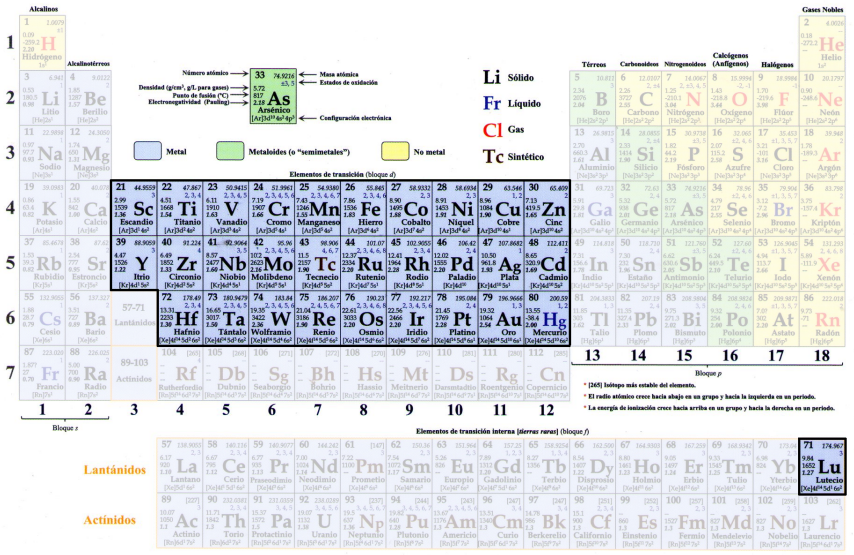

Figura 10: Elementos de la tabla periódica que vamos a estudiar.

Los elementos seleccionados son todos metales que poseen electrones en el orbital d. El orbital d es un nivel energético que puede albergar hasta 10 electrones. Los elementos en estudio son los correspondientes al nivel energético n=3,4 y 5, y todos los posibles llenados del orbital d, desde d=1, hasta d=10.

Gracias al programa `SIESTA`, podemos realizar una simulación de cada elemento en la que podemos obtenemos la energía y su densidad de estados. El caso de estudiar las estructuras con la misma constante de red, puede ser de gran ayuda en el estudio de la influencia de la configuración electrónica en la DOS. De esta forma, obtenemos para cada una de las estructuras, dos matrices, la matriz de energías y la matriz DOS. La matriz de energías es una matriz perteneciente a M_{30x800} $(\mathbb{R})$ en la que cada fila representa un elemento diferente,

y cada columna, los valores de de la energía en 800 puntos. La matriz DOS es una matriz perteneciente a M_{30x800} $(\mathbb{R})$ al igual que la matriz de energías, solo que, esta vez, los valores corresponden a la DOS de cada uno de los 30 elementos diferentes.

$$DOS = [d^1, ..., d^{30}]. \tag{15}$$

En este caso cada d^i es un vector fila que contiene los 800 puntos de la DOS del elemento i. Si representamos la DOS en función de la energía para el elemento Escandio (Sc), por ejemplo, obtenemos la figura (11).

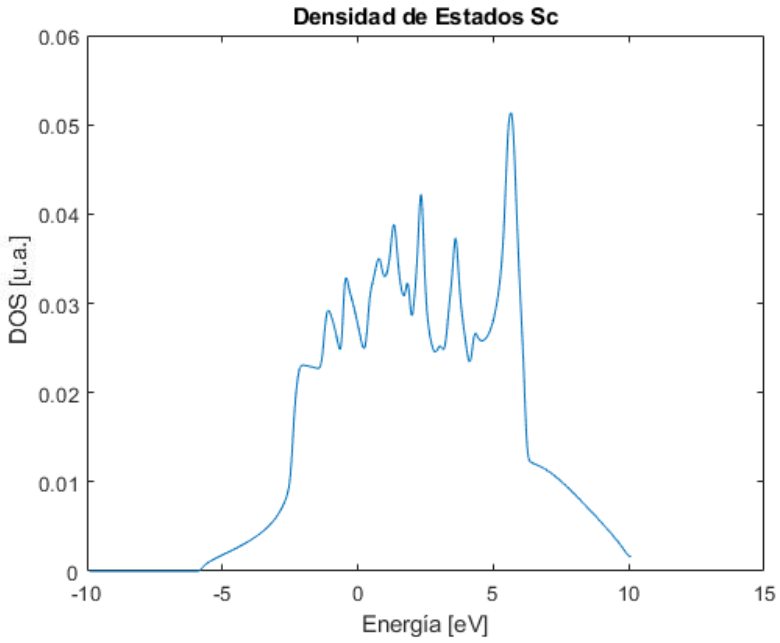

Figura 11: DOS del elemento Escandio para cada valor de la energía dada en eV.

En el caso de materiales cristalinos, se consideran estructuras infinitas y debido al uso de condiciones de contorno periódicas, las deltas de Dirac presentes en materiales finitos (como el ejemplo mostrado del Hidrógeno), pasan a una curva continua fruto de la unión de todas las curvas que intervienen. La curva DOS está normalizada de forma que realizando la integral desde el menos infinito hasta la energía de Fermi, en 0:

$$\int_{-10}^{0} DOS_{Sc}(E)dE = 3, \tag{16}$$

la integral nos indica el número de electrones en la capa de valencia:

$$Sc\ [Ar]\ 3d^1 4s^2. \tag{17}$$

Luego tenemos tres electrones de valencia, tal y como podemos comprobar de la ecuación (16).

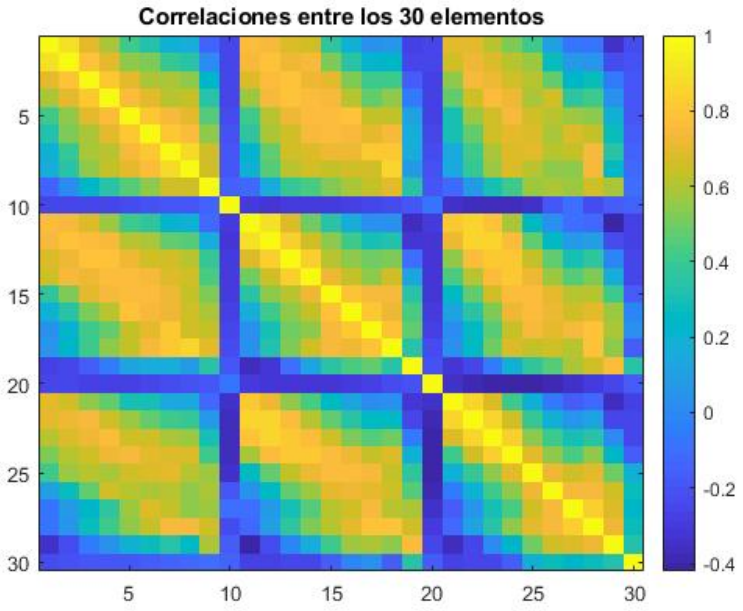

Figura 12: Correlaciones entre las curvas DOS para los 30 elementos.

La imagen (12), representa las correlaciones existentes entre las curvas DOS para cada uno de los 30 elementos de estudio. El gráfico de color muestra en amarillo, los elementos de correlación 1. Cabe destacar cómo los elementos 10, 20 y 30, correspondientes al Cinc, Cadmio y Mercurio respectivamente, tienen una correlación nula o incluso negativa en algunos casos. Estos elementos son los últimos elementos de cada fila y tienen todas sus subcapas electrónicas llenas. En concreto, la figura corresponde a la estructura FCC calculada con la constante de red más estable para cada elemento. El resultado en las 7 estructuras restantes es similar.

n el cambio en el

Junto con las curvas DOS, tenemos paila, e indicara cada elemento los valores correspondientes a:

- Número atómico: El número atómico, o Z, es el número de protones que tiene el átomo. En el caso de un átomo neutro, de carga total cero, el número de electrones coincide con Z.

- Electrones de valencia: Los electrones pertenecientes a capas no completas se denominan electrones de valencia, y están presentes en procesos de enlace e interacción con otros átomos.

- Número cuántico de la capa de valencia: Indica el nivel energético y determina las filas en la tabla periódica, en donde cada fila se refiera al número principal n. En nuestro caso tendremos en cuenta elementos de la cuarta, quinta y sexta fila, correspondientes a n=4, 5 y 6 respectivamente.

- Electronegatividad de Pauling: La electronegatividad es la capacidad que tiene un átomo de atraer electrones hacia si mismo. Los átomos con mayor electronegatividad, intentarán atraer los electrones de los átomos y alejarlos de los átomos a los que se una. Cuantos más protones tenga un átomo, mayor será la atracción del núcleo con carga positiva hacia los electrones de carga negativa. Si los electrones están en orbitales muy grandes, la distancia con el núcleo y el apantallamiento del resto de electrones hace que la atracción con el núcleo sea menor. Por eso la electronegatividad aumenta con el número atómico, y disminuye cuando el radio atómico aumenta. Si avanzamos de izquierda a derecha en la misma fila de la tabla periódica, correspondiente a elementos con electrones de valencia en el mismo nivel de energía (n), el número de protones aumenta, disminuye el radio atómico debido a la mayor atracción y la electronegatividad aumenta. Si avanzaos de abajo a arriba en la misma columna, aunque el número de protones disminuye, el radio atómico también disminuye haciendo que la atracción de los electrones sea más potente. Por eso los elementos situados en la tabla arriba a la derecha son los más electronegativos, mientras que los elementos abajo a la izquierda son los menos electronegativos. Como los gases nobles, que tienen todas sus capas llenas, no atraen electrones, el Flúor es el elemento más electronegativo. La electronegatividad explica procesos como la cesión de electrones poco electronegativos, como el Sodio, a elementos más electronegativos como el Cloro en el enlace iónico NaCl.

Linus Pauling (1901 - 1994) fue el primero en estudiar la electronegatividad. El método de Pauling consiste en mirar el enlace de una molécula formada por dos átomos, X e Y. La energía de enlace teórica es el promedio de la energía de enlace X-X y la energía de enlace Y-Y. Después comparamos la energía de enlace teórica con la energía de enlace experimental medida sobre el enlace X-Y;

$$\Delta E = (X-Y)_{\text{experimental}} - (X-Y)_{\text{teórico}} \tag{18}$$

Si ΔE es cero, significa que la electronegatividad de X e Y coincide, si no es cero, un átomo intentará captar electrones convirtiendo el sistema en una molécula polar[5], siendo parcialmente negativa. Pauling calculó los valores de la electronegatividad a partir del Flúor, al que le asignó un valor de 4.0, siendo el resto de valores relativos a éste.

- Radio del orbital s: Debido al comportamiento cuántico del electrón, no podemos asegurar una trayectoria que describa su movimiento al igual que lo hace la Tierra alrededor del Sol. Sin embargo el orbital nos indica una región de mayor probabilidad en la que podemos encontrar el electrón. El radio s nos proporciona el radio del orbital s de cada elemento. Debido a la diferencia de protones en el núcleo, el efecto atractivo de diferencia de cargas con los electrones hace que estos estén más cerca del núcleo, reduciendo el radio. Los niveles energéticos más altos, también están más alejados del núcleo, luego cabría esperar que el elemento de nivel energético más grande, y menor número de protones, sea el que tenga un radio del orbital s mayor, lo que hace que cada elemento tenga un radio independiente. Si consultamos los datos vemos que efectivamente el Lutecio es el elemento de mayor radio con un valor de 1.35 Å.

- Radio del orbital p: Distancia más probable de encontrar el electrón en el orbital p. El mismo razonamiento que en el radio s se cumple, siendo el Lutecio el de mayor tamaño, esta vez con 1.66 Å.

[5]Una molécula polar es aquella que tiene un momento dipolar, es decir, que el reparto de carga no es uniforme y posee una parte con carga negativa y otra positiva.

- Radio del orbital d: Distancia más probable de encontrar el electrón en el orbital d, análogo al s y el p con un radio máximo para el Lutecio de 0.68 Å.

- Número de Mendeleev (M.N): El número de Mendeleev corresponde a un valor natural que se asigna a cada elemento de acuerdo a la clasificación realizada por Dimitri Mendeleev (1834 - 1907). Es un valor análogo al número atómico que ordena la tabla periódica actual, pero en este caso no está relacionado con el número de protones. La clasificación realizada por Mendeleev estaba fundamentada en la idea de que las propiedades de los elementos eran una función periódica de sus masas atómicas.

- Escala química: *Chemical scale* o escala química (E.Q.) es una magnitud adimensional que nos da información de la estructura cristalina de compuestos binarios, aquellos formados por dos especies atómicas. La escala química admite diferentes definiciones que consideran propiedades atómicas como la electronegatividad. En los metales de transición, las variables principales que definen la escala química son los electrones de valencia en los orbitales s, p y d

- Constante de red: La constante de red indica la distancia entre los primeros vecinos de cada estructura. Los datos de esta variable corresponden a la constante de red más estable en la que el elemento se encuentra en la naturaleza, por lo que es diferente de cada estructura.

- Energía de Fermi: La energía de Fermi fue introducida en la definición de la desanidad de estados. También podríamos interpretarla como la diferencia energética entre el estado más alto y más bajo de un sistema cuántico a temperatura 0 K. Normalmente es la diferencia de energías cinéticas en un gas de Fermi[6].

[6]Modelo físico que considera un sistema ideal de fermiones libres que no interaccionan entre si

3. Algoritmo de Clusterización.

Uno de los problemas habituales que encontramos en el desarrollo de la investigación, es el agrupamiento o clusterización de elementos. Es por ello que se plantea un problema a la hora de realizar los grupos. Nuestro objetivo es eliminar en la medida que sea posible el factor subjetivo que podemos tener en este tipo de situaciones, por ello el planteamiento de un algoritmo que realice la clasificación puede ser de gran interés. Si bien es verdad que es complicado que la clasificación sea la adecuada para análisis a posteriori de los datos, o situaciones individuales en problemas diferentes, esta herramienta puede servir de apoyo en la toma de elecciones dando una razón matemática más allá de la subjetividad que nosotros le podemos dar.

3.1. k-means ++.

El método k-means tiene como objetivo la agrupación de elementos minimizando las distancias de los elementos del mismo grupo. Se trata de un algoritmo simple y veloz, que lo convierte en una técnica muy utilizada a pesar de no ser una de las más precisas.

El problema k-means, parte de un numero entero k, y un conjunto de n puntos X de dimensión $\mathbb{R}^d$. Hay que encontrar k centros denotados por el conjunto C, que minimicen la función potencial ϕ, que representa la distancia de cada punto a su centro más cercano, siendo la distancia considerada la euclídea d-dimensional.

$$\phi = \sum_{x \in X} \min_{c \in C} \|x - c\|^2 \tag{19}$$

3.1.1. Algoritmo k-means.

El algoritmo k-means es el algoritmo más sencillo y rápido de clusterización, pero también es el menos eficaz. Su descripción es la siguiente:

1. Primero seleccionamos k centros iniciales, denotados por el conjunto C=$\{c_1, c_2, ...c_k\}$. La selección se suele realizar de manera arbitraria de forma uniforme en X.

2. Para cada i $\in \{1, ..., k\}$ denotamos C_j como el conjunto de puntos de X que están más cerca de c_i que de cualquier c_j $\forall$ $j \neq i$. En caso de empate, la elección del centro puede

ser arbitraria.

3. Para cada $\{1, ..., k\}$, denotamos por c_i al centro de masas de todos los puntos de conjunto C_i.

$$c_i = \frac{1}{|C_i|} \sum_{x \epsilon C_i} x \tag{20}$$

4. Se repten los pasos 2 y 3 hasta que el conjunto C no cambia.

3.1.2. k-means++.

Como mejora del algoritmo k-means, k-means++ propone un modo específico a la hora de tomar los centros:

1. Escogemos un centro c_1 de forma aleatoria y uniforme de X.

2. Escogemos un nuevo centro c_i, tomando $x \in X$ con probabilidad:

$$P_x = \frac{D(x)^2}{\sum_{x \epsilon X} D(x)^2}, \tag{21}$$

   donde $D(x)$ es la distancia del punto x al centro centro más cercano de los ya definidos. D^2 es denominado peso.

3. Repetimos el paso 2 hasta haber obtenido exactamente k centros.

4. Repetimos los pasos 2-4 del algoritmo k-means.

3.2. Optimización del número de grupos.

El algoritmo k-means++ nos proporciona los grupos óptimos para un k dado, pero la elección de ese k no es trivial. Por ello se plantea el siguiente algoritmo de optimización del número de clusters k:

1. Partimos de una muestra de n elementos que queremos clasificar y denotamos N, por el número de grupos en los que se reparten los n elementos. Es evidente que si N=1, el problema carece de interés, por lo que establecemos un valor mínimo de grupos, o

clusters, de N=2. Un razonamiento análogo surge al considerar N=n, ya que si todos los elementos son únicos en cada grupo, el problema carece de interés. De esta forma el valor máximo de N será al menos n-1. Una vez establecidos el número de grupos mínimo y el número de grupos máximo, se consideran todas las posibilidades intermedias, siendo estas N=2,3,,,n-2,n-1. El rango de valores se puede restringir solo a los valores centrales que pueden resultar de más interés. Para cada N diferente, se realiza una clasificación de los elementos gracias al comando $kmeans(X, k)$ donde X es una matriz de dimensión al menos 2x2 y k especifica el número de grupos, utilizando el algoritmo K-means++.

2. Para cada N diferente, calculamos el centro de gravedad de cada uno de los N grupos. El centro de gravedad para el grupo i-ésimo formado por m elementos de dimensión (x,y,...,t) es:

$$C^i = \sqrt{\left(\frac{x_1^i + ... + x_m^i}{m}\right)^2 + \left(\frac{y_1^i + ... + y_m^i}{m}\right)^2 + ... + \left(\frac{t_1^i + ... + t_m^i}{m}\right)^2} \tag{22}$$

se trata de una norma euclídea de los los promedios de cada elemento por componente.

3. Calculamos la desviación típica del conjunto de distancias de cada elemento del grupo a su centro de gravedad σ_N^i. Al igual que en la elección del centro, la distancia se tomará como la norma euclídea de la diferencia del centro de gravedad y cada elemento

4. Calculamos la distancia entre los centros de gravedad de los N grupos: para cada grupo ij de los N posibles, d_N^{ij}.

5. Para cada par de dos grupos ij, se calcula el ratio:

$$F_{ij} = \frac{d_N^{ij}}{\frac{\sigma_N^i + \sigma_N^j}{2}}$$

6. En el caso en el que un grupo tenga un único elemento, la desviación típica de las distancias de los elementos de dicho grupo será cero. Lo mismo ocurre en un grupo en el que solo existan dos elementos. Como la distancia entre los dos elementos al centro de

gravedad es la misma, esto nos da dos valores degenerados, que tienen una desviación típica igual a cero. Por ello, en el caso en el que un grupo tenga una desviación típica igual a cero, se tomará como desviación del grupo, la máxima desviación típica entre las desviaciones de todos los grupos, considerando todos los posibles grupos para cada N. Es decir, si σ_N^i, la sigma del grupo i, es igual a 0, entonces:

$$\sigma_N^i = \text{máx}\{\sigma_N^j\} \quad \forall j = 1, .., N \quad \forall N = 2, ..., n-1. \tag{23}$$

7. El objetivo principal es maximizar la distancias entre grupos, minimizando la distancia entre los elementos de dicho grupo. Nos interesa entonces que el ratio F_{ij} sea lo más grande posible. Para valorar todos los ratios calculados en cada N diferente, tomamos su valor medio. Para un grupo de s distancias.

$$F_{Tot} = \Sigma_{i<j}^{s} \frac{F_{ij}}{s} \tag{24}$$

El N que genere el mayor F_{Tot}, será el óptimo. Podemos ver que en el caso en el que se considerase un caso N con dos grupos de desviaciones típicas nulas, el ratio F tomaría un valor infinito, provocando que dicho grupo fuese tomado como óptimo. Como se trata de un algoritmo de agrupación, se ha optado por poner el caso menos favorable para grupos unipuntuales o de solo dos elementos, dando la mayor desviación de sus elemento y disminuyendo F.

Otra forma de resolver este problema sería exigiendo un mínimo de tres elementos por grupo, pero para elementos muy aislados, esto podría suponer un desplazamiento de los centros de gravedad provocando un error mayor y una pérdida de información en la clusterización.

Por último también se plantea un análisis previo aislando del proceso de clusterización aquellos elementos muy dispares con el resto, pero de nuevo entra en juego una medida subjetiva en el análisis previo que queremos eliminar, por ello se ha tomado la primera

opción como la más válida.

3.3. Análisis del algoritmo de clusterización.

El objetivo del análisis de clusterización era eliminar el criterio subjetivo en la clasificación de elementos. Una vez desarrollado y probado, analizamos si el resultado final ha sido bueno y proponemos posibles mejoras que puedan desarrollarse en el futuro.

El criterio final al realizar una clasificación, siempre puede verse afectado por otro tipo de variables externas que el algoritmo no pueda controlar. Por ejemplo este algoritmo no permite establecer un número medio de elementos de cada grupo para formar así una división homogénea en el número de individuos. El tipo de problemas de clasificación con los que nos hemos encontrado, principalmente clasificaciones de los elementos químicos de la base de datos, no tienen por qué formar grupos homogéneos. Han sido estos problemas los que han motivado el diseño de este algoritmo, con el objetivo de dar un mayor rigor a la investigación.

Otra posible mejora, sería la de realizar un análisis previo en la que se determine la condición óptima para aplicar el algoritmo. Nuestra base de datos presenta unos valores lo suficientemente cercanos como para que ningún elemento, o grupo de ellos, tuviese que tener un tratamiento especial. En principio, cualquier elemento puede estar relacionado. Sin embargo esto no tiene por qué ser así. El algoritmo podría aplicar un análisis previo, que dividiese la base de datos en subgrupos, los cuales a su vez se les podría aplicar de nuevo el algoritmo. Así podría conseguirse grupos de distintos niveles de clasificación, dependiendo del número de veces que se ha aplicado el algoritmo.

Una vez comentado esto, y centrándonos en la base de datos en la que trabajamos durante esta investigación, pasamos a analizar su eficiencia. El algoritmo presenta el comportamiento esperado en la mayoría de casos. Para grupos que presentan valores muy claros, el algoritmo da el valor esperado óptimo. Sin embargo, en nubes de puntos cercanos, el número de grupos óptimo puede variar. Esto es debido al uso del algoritmo k-means++. Como se ha explicado,

la presencia de dos casos igual de probables, hace que la selección de centros pueda no ser la misma. Esto provoca que no siempre se formen los mismos grupos con los mismos elementos. El algoritmo aquí propuesto parte de k-means++, por lo que corregir esto supone corregir, o cambiar, el algoritmo con el que se hacen los grupos y que luego vamos a optimizar.

La diferencia del número de grupos óptimo puede variar en una, o dos unidades en casos más extremos. Teniendo esto en cuenta, se ha procurado escoger el número de grupos óptimo más habitual. Sin embargo, al realizar varias clasificaciones de los elementos en la que cada clasificación depende de na variable e interesa que el número de grupos sea constante en cada clasificación, se ha exigido un número de grupos medio, aunque este pueda no ser el óptimo.

Como conclusión final, el algoritmo diseñado ha sido de gran ayuda a la hora de establecer clasificaciones, pero no ha sido capaz de eliminar totalmente la presencia del investigador. Es una herramienta que puede ser muy útil como punto de partida que se ha adaptado bastante bien con nuestro problema concreto. Como mejora se propone cambiar el algoritmo de clasificación inicial, realizar un análisis previo de los datos iniciales e introducir nuevas variables de control como el número de elementos máximo, o mínimo, por grupo, o establecer una distancia máxima para que un elemento pertenezca a un grupo concreto.

4. Predicción de la Densidad de Estados (DOS).

4.1. Base PCA.

Principal Component Analysis, o PCA, es una técnica que mediante transformaciones ortogonales sobre un conjunto de observaciones que pueden estar correlacionados, obtiene un nuevo conjunto de variables no correlacionadas linealmente, que toman el nombre de variables principales. El objetivo de una base PCA es minimizar la dimensión de una muestra, maximizando la información. Vamos a aplicar una PCA sobre la matriz DOS. Lo primero que hacemos es centrar la matriz. Para ello calculamos el valor medio de cada una de las columnas.

$$\mu^j = \sum_{i=1}^{30} \frac{d_{ij}}{30} \ \forall j = 1, .,800. \tag{25}$$

Si a cada una de las 30 filas de la matriz DOS le restamos el vector que contiene las medias de cada columna, obtenemos la matriz DOS centrada, es decir, $CDOS$.

$$CDOS = DOS - \boldsymbol{\mu} \ \in M_{30\times800}, \tag{26}$$

donde $\boldsymbol{\mu} \in M_{30x800}$ y $\mu_{ij} = \mu^j$.

Una vez centrada, multiplicamos la matriz transpuesta de $CDOS$ por $CDOS$, obtenido así una nueva matriz C $\in M_{30\times30}(\mathbb{R})$.

$$CDOS^t \cdot CDOS = C \ \in M_{30\times30}(\mathbb{R}). \tag{27}$$

Descomponemos C en un producto $PDP^{\,t}$, en el que P y $D \in M_{30\times30}(\mathbb{R})$, y D es una matriz diagonal, de la cual obtenemos sus valores propios $\{ \lambda_1, ..., \lambda_{30} \}$, con $\lambda_i \geq \lambda_j \ \forall$ i $\geq$ j, con i, j$\in\{1,...,30\}$. Debido al centrado de los datos realizado, perdemos una dimensión, por lo que λ_{30}=0.

$$C = P \cdot \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_r \end{bmatrix} \cdot P^t \tag{28}$$

Una vez calculados los valores propios, definimos la energía de índice q, E_q como:

$$E_q = \frac{\sum_{k=1}^{q} \lambda_k}{\sum_{k=1}^{30} \lambda_k} \cdot 100 \tag{29}$$

Está claro que E_q es un porcentaje que nos ayudará a reducir la dimensión. Par ello, buscamos el índice q necesario para que E_q sea mayor que el 99 %.

Creamos la base PCA, que se trata de una matriz $\in M_{800\times30}(\mathbb{R})$, en la que cada columna está dividida por la norma euclídea de los elementos de la matriz resultado del producto de $CDOS^{\,t} \cdot P$.

$$PCA_{ij} = \sum_{r=1}^{30} CDOS_{ir}^{t} \cdot P_{ir} \tag{30}$$

$$PCA_{ij} = \frac{PCA_{ij}}{\| \sum_{j=1}^{30} PCA_{ij}^{2} \|^{\frac{1}{2}}} \tag{31}$$

La matriz PCA es una matriz auxiliar que podría verse como la matriz PCA antes de aplicarse la normalización.

De ésta última matriz PCA reducimos la dimensión de columnas al índice q, obteniendo una matriz $PCAR \in M_{800\times q}(\mathbb{R})$ en la que las columnas de $PCAR$ coinciden con las primeras q columnas de PCA.

$$PCAR_{ij} = PCA_{ij} \;\; \forall i \in \{1, ..., 30\}, j \in \{1, ..., q\}. \tag{32}$$

Por último, para obtener las coordenadas de PCA buscadas, multiplicamos:

$$CPCA = DOSC \cdot PCAR \quad \in M_{30\times q}(\mathbb{R}). \tag{33}$$

Si representamos las dos primeras columnas de $CPCA$, $PCA1$ y $PCA2$, respectivamente, que corresponden a las coordenadas que representan máxima variación, obtenemos, por ejemplo, para la estructura FCC formada por la constante de red más estable para cada elemento:

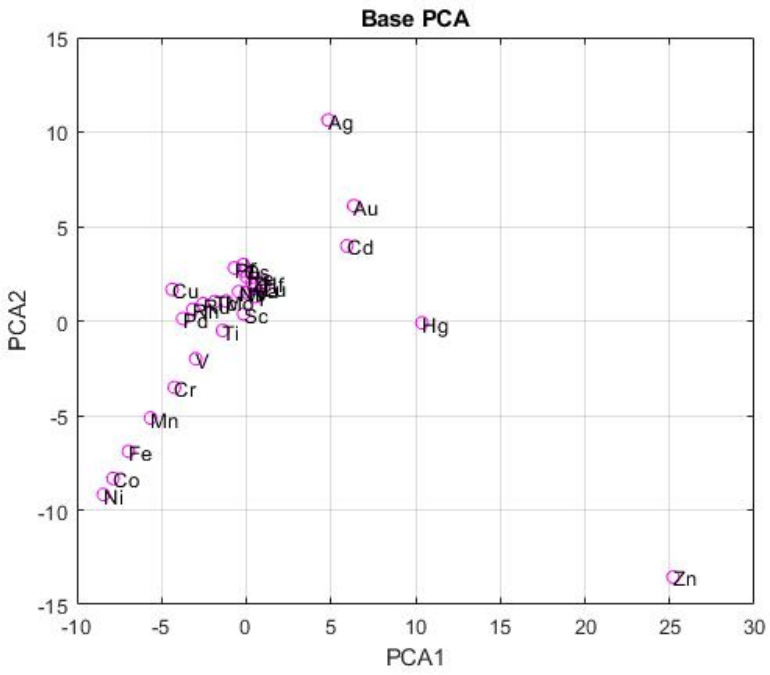

Figura 13: Representación gráfica de $PCA1$ y $PCA2$ para los elementos de estructura FCC.

Podemos reconstruir las curvas DOS suavizadas, es decir, a través de la reducción de dimensión,

$$DOSR = PCAR \cdot CPCA^t + \boldsymbol{\mu} \tag{34}$$

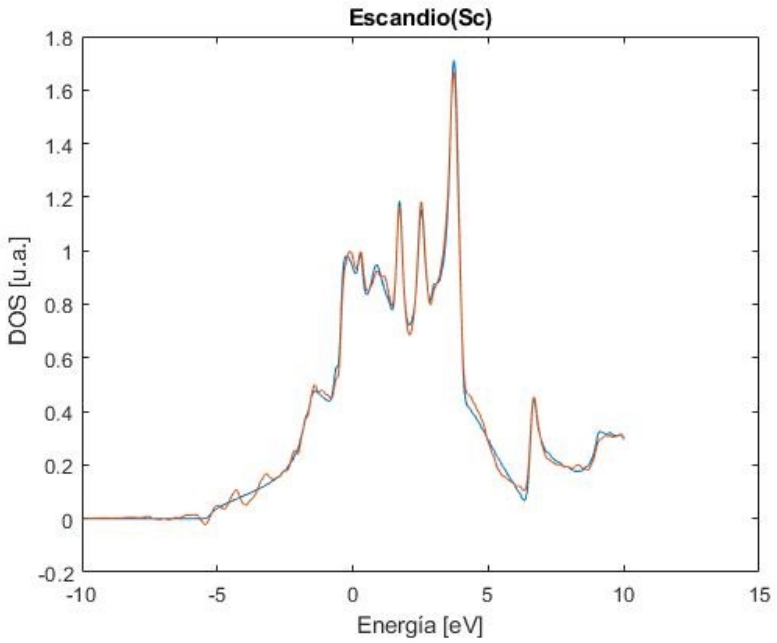

Figura 14: Comparación entre la curva original, en azul, y la curva suavizada reconstruida en rojo.

4.2. Análisis de las curvas.

Nos enfocamos ahora en el estudio y clasificación de las DOS obtenidas. Podemos agrupar a los 30 elementos en 5 grupos diferentes, analizando solamente las dos primeras componentes de la base PCA. Para el caso concreto de la estructura FCC, los grupos son los obtenidos en la figura(15)

De cada elemento obtendremos información extraída directamente de las curvas DOS. En cada curva vamos a escoger unos puntos de interés, que equivalen al intervalo que encierra mayor área, como muestra la figura (16). Dentro de esta selección, los valores de interés son los siguientes:

- El valor de la energía del primer punto de la curva seleccionada (P. inicial DOS).
- El valor de la energía del último punto de la curva seleccionada (P. final DOS).
- La anchura de la curva seleccionada.
- El valor en la DOS del máximo que alcanza la curva (Máximo DOS).

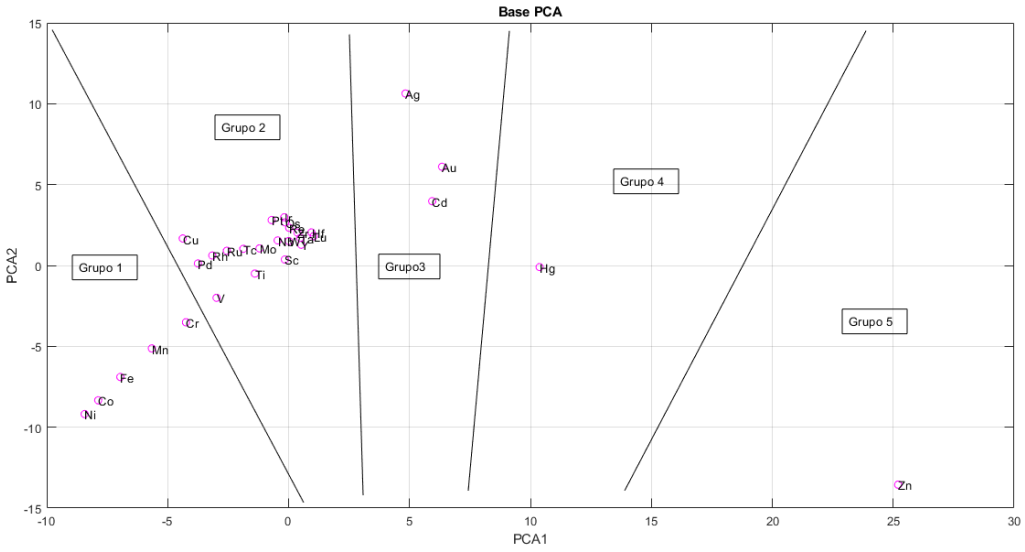

Figura 15: Representación gráfica de $PCA1$ y $PCA2$ para los elementos de estructura FCC.

- El valor en la energía del máximo que alcanza la curva (E. Max DOS).
- El valor en la DOS del mínimo que alcanza la curva seleccionada (mínimo DOS).
- La diferencia entre el máximo y mínimo mencionados anteriormente (Dif Max min DOS).
- El número de puntos críticos que tiene la curva seleccionada (Pts Críticos).

El objetivo de este análisis consiste en realizar un análisis de atributos sobre las DOS, para aumentar los datos iniciales de cada elemento, pasando de 11 datos iniciales, a un total de 19.

Para la selección de los atributos de las curvas se ha realizado el siguiente proceso. Nos centramos en obtener cuál va a ser el primer punto de la curva, y quién va a ser el último en el eje de las energías. Después, basta con tomar los valores intermedios para completarla. Primero seleccionamos el máximo de la curva DOS, y a partir de él, seleccionamos una tolerancia para que se tomen los puntos de las energías que corresponden a una altura superior que 100-α % del máximo, tomando un valor pequeño, en este caso se seleccionó α=10. De esta forma hemos conseguido una curva de puntos en un entorno del máximo. Ahora sumamos el valor de los puntos en la curva DOS, y lo dividimos por la suma total de

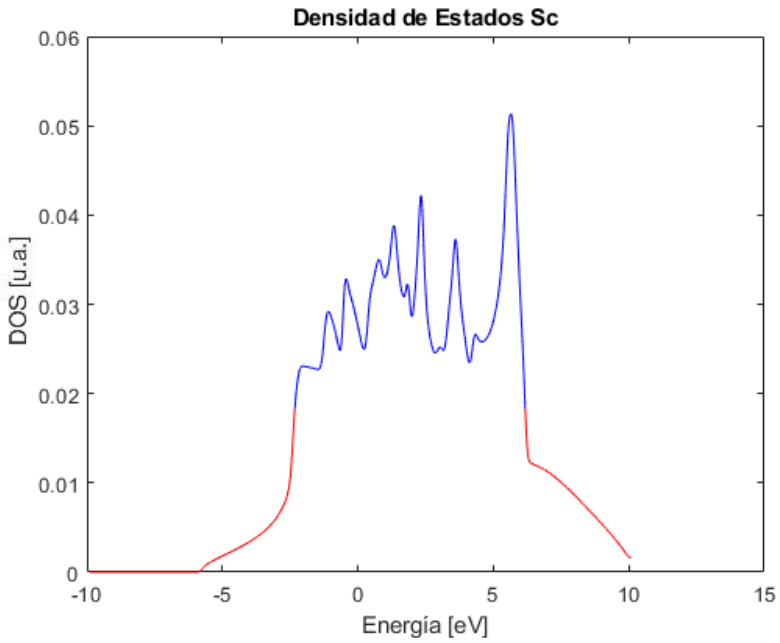

Figura 16: En Azul la curva seleccionada para el análisis. En rojo la curva total.

la DOS así obtenemos un valor entre cero y uno, que representa una fracción seleccionada de la curva total. A partir de aquí, si aumentamos el valor de α iterativamente, vamos reduciendo la tolerancia, y aumentando el número de puntos de la curva que queremos seleccionar. En cada grupo de los obtenidos en la base PCA, seleccionamos el valor del cociente de la curva seleccionada y la total que más nos interesa.

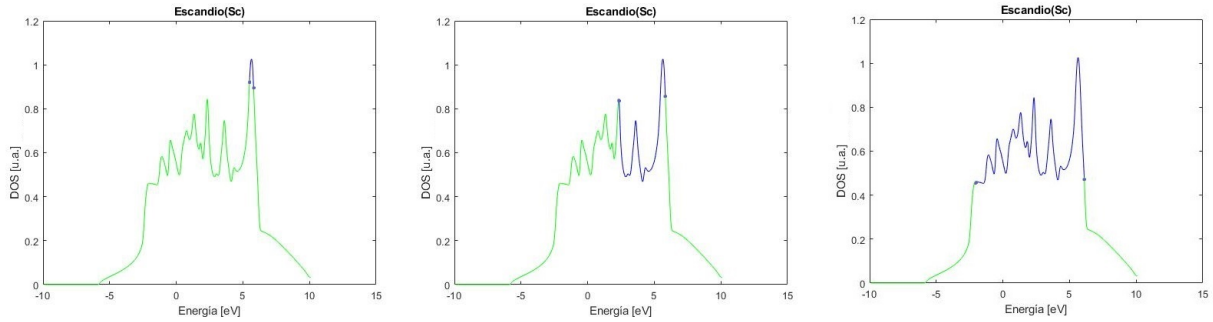

Figura 17: Esquema del proceso de selección de la curva deseada en tres procesos diferentes. De izquierda a derecha se observa cómo al aumentar la tolerancia, aumenta el número de datos seleccionados.

Para la selección de puntos críticos, tomamos como punto de partida los valores que toma

la DOS en la curva seleccionada. La selección se realizará a través del signo de la derivada. En los puntos en los que se produzca un cambio de signo, se habrá alcanzado un punto crítico.

La manera que tenemos de comprobar cómo de buena es la selección de las curvas es a través de los radios de los orbitales. Ya hemos visto que existe una relación directa entre los electrones de valencia y la DOS, luego cabría esperar un valor de correlación muy alto entre las anchuras seleccionadas, y los radios atómicos. Como ejemplo los valores obtenidos analizando las covarianzas son para las estructuras SC, FCC, BCC y HCP calculadas con la misma constante de red para cada elemento son:

Estructura	Radio s	Radio p	Radio d
SC	0.8287	0.8306	0.7135
FCC	0.9041	0.8998	0.8204
BCC	0.9415	0.9383	0.8531
HCP	0.9022	0.8897	0.8215

Cuadro 1: índices de correlación para el radio y la anchura de las DOS para cada estructura en estructuras formadas por la constante de red más estable en la naturaleza.

Los valores cercanos a 1 indican que, efectivamente, existe una relación entre ambas variables que indica que la elección de la curva se ha realizado correctamente.

Cabe destacar que, debido a la diferencia que presentan las DOS en la estructura SC, la selección de la curva de interés es más compleja, y tiene un peor resultado. Aún así, la correlación es alta, y nos servirá de guía en el siguiente proceso.

4.3. Predicción de DOS.

Pasamos ahora a la predicción de la curvas DOS. Esta predicción utiliza tantos las matices obtenidas en el análisis de la base PCA, como todos los atributos obtenidos hasta ahora. La idea de la estimación es obtener un conjunto de pesos utilizando un ajuste de mínimos cuadrados. Dichos pesos se combinan con los atributos obtenidos para así llegar al valor buscado.

El algoritmo utilizado para la predicción de las curvas es el siguiente:

- Identificamos la matriz de atributos como A_j, donde j representa las estructura. De esta forma contamos con 8 matrices de atributos formadas por 30 filas que representan a los elementos y 19 columnas que representan los diferentes atributos. Fijamos ahora una matriz A_j, correspondiente a una estructura concreta.

- Establecemos el número de predicciones simultaneas que vamos a realizar, $n \in \mathbb{N}$. Escogemos, aleatoriamente, los elementos que van a participar en la predicción, $\mathbf{m}$. Se trata de un vector en el que cada componente identifica a un elemento, siendo el 1 es escando, etc. Así una predicción donde $\mathbf{m}$=(1,15,26), sería una predicción que utiliza los elementos 1, que corresponde al Escandio, el 15, correspondiente al Tecnecio y el 26, el Osmio. El valor de $\mathbf{m}$ varía para cada una de las n predicciones simultaneas.

- Construimos la matriz de pesos $P^{\mathbf{m}}$ de cada predicción como:

$$P^{\mathbf{m}} = (A_j^{\mathbf{m}})^{\dagger} * CPCA_j^{\mathbf{m}} \qquad \forall i = 1, .., n. \tag{35}$$

  La matriz $CPCA_j$ se obtuvo en la base PCA para cada estructura. como cada fila representaba un elemento, $CPCA_j^{\mathbf{m}}$ es una matriz en la que se han seleccionada los elementos determinados por $\mathbf{m}$.

- Para realizar la reconstrucción introducimos la matriz de coordenadas $PCAR$ obtenida en la base PCA:

$$R_j = PCAR \cdot (A_j \cdot P^m)' + \boldsymbol{\mu} \in M_{30\times 800}(\mathbb{R}). \tag{36}$$

  La matriz R_j es una matriz en la que cada fila representa un elemento, y las columnas son las predicciones de las curvas DOS.

- Cada R_j representa una predicción simultánea, luego realizando una media por cada elemento de matriz, obtendríamos la predicción final, es decir, Denotamos por RF, a

la reconstrucción final, entonces el elemento de matriz de la fila k, columna l de RF es:

$$RF_{kl} = \frac{\sum_{j=1}^{n} R_{kl}^{j}}{n}. \tag{37}$$

La matriz resultante contiene las curvas DOS predichas para los 30 elementos, y es independiente del valor de **m**.

4.3.1. Análisis de los resultados.

El algoritmo mostrado anteriormente tiene como objetivo determinar la curva DOS de un elemento a través de su $finger\ print$, es decir, el conjunto de atributos que lo describen. Para ello se han considerado varias predicciones simultáneas, que utilizan elementos diferentes, y atributos diferentes en cada predicción.

Como ejemplo de predicción mostramos la reconstrucción que se a ha realizado sobre el Escandio, en una estructura FCC considerando la constante de red más estable, en azul la curva teórica y en rojo la predicción:

La figura (18) muestra la reconstrucción utilizando 5 predicciones simultáneas y un número mínimo de elementos de 10. La predicción no ha sido muy buena por lo que probamos a incrementar el número de elementos que intervienen en la predicción y el número de predicciones simultaneas.

Realizamos ahora una reconstrucción de la curva DOS con 10 predicciones simultáneas y un mínimo de 15 elementos. En azul la curva teórica y en rojo la predicción.

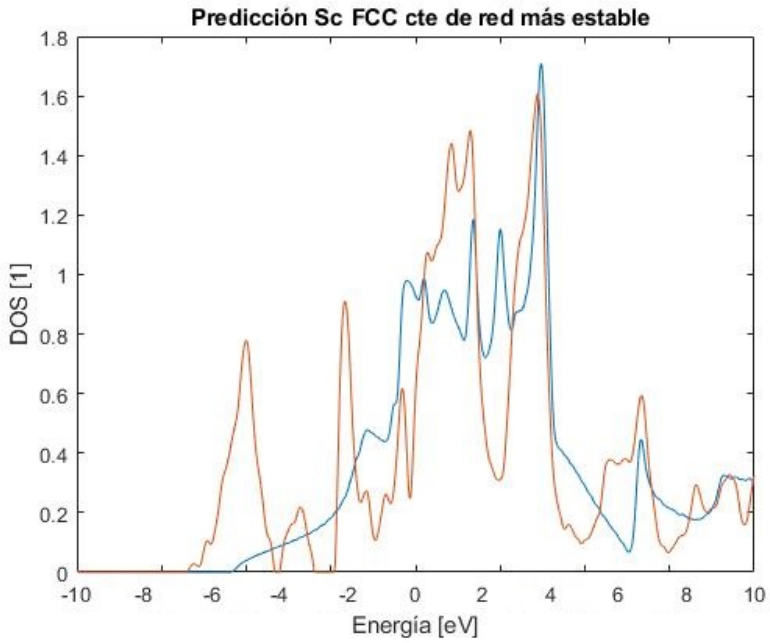

Figura 18: Primera predicción del Escandio.

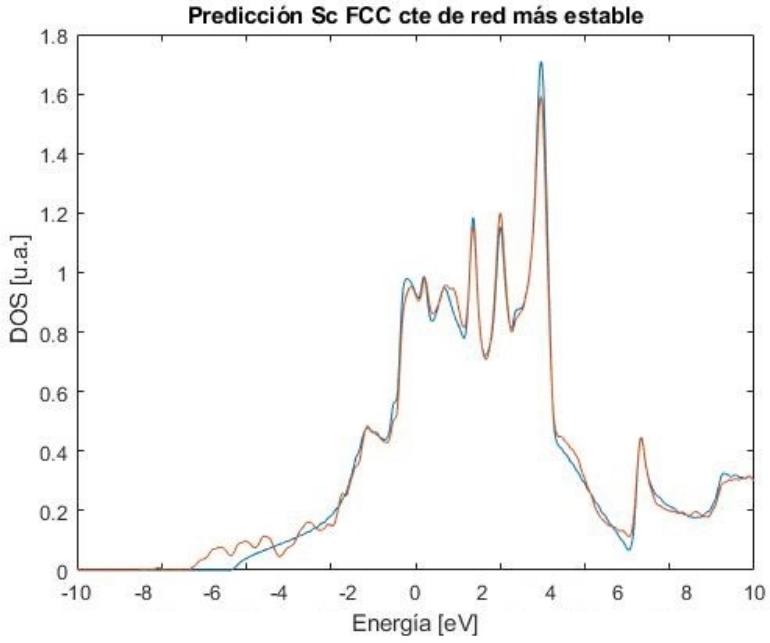

Figura 19: Segunda predicción del Escandio.

Como puede comprobarse visualmente, el ajuste es excelente, sobretodo en la parte central, que es la de mayor interés. La posición y altura de los puntos críticos tiene muy buen

comportamiento. Calculamos su error como:

$$\text{error} = \frac{\|DOS_{Sc} - RF_{Sc}\|_\infty}{\|DOS_{Sc}\|_\infty} \cdot 100. \tag{38}$$

El error obtenido es del 10.2278 %. Es un error pequeño, causado sobre todo en el extremo de menor energía.

El resultado obtenido es muy satisfactorio. El aumento de la precisión de la estimación mejora considerablemente al aumentar el número de predicciones simultaneas y el número mínimo de elementos involucrados. El número de predicciones simultáneas no es un número muy grande, y el número mínimo de elementos es de tan solo la mitad. Aumentar las predicciones simultaneas no supone un coste muy grande ya que el algoritmo es rápido.

El comportamiento en el resto de estructuras es exactamente el mismo:

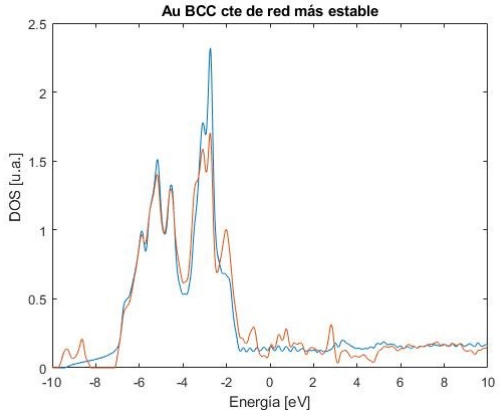

Figura 20: Predicción Oro estructura BCC con la cte de red más estable

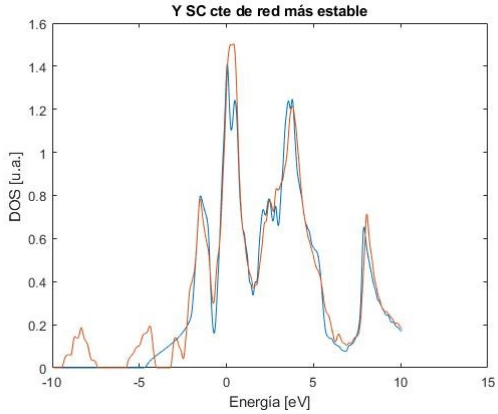

Figura 21: Predicción Itrio estructura SC con la cte de red más estable

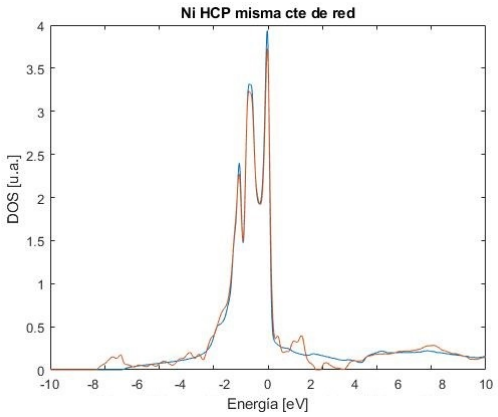

Figura 22: Predicción Níquel estructura HCP con la misma cte de red para cada elemento.

Los errores cometidos son:

Elemento	Error
Oro	26.7818 %
Níquel	8.8989 %
Itrio	20.1394 %

Cuadro 2: Tabla de errores.

Aunque el error cometido en el Oro y el Itrio son grandes, es más interesante la posición de los máximos y su altura, los cuales sí que están bien ajustados.

Los resultados finales han sido bastante buenos. La predicción de las curvas DOS ha tenido muy buen comportamiento en todas las estructuras independientemente de la constante de red escogida. Aún con pocas predicciones simultáneas los valores obtenidos han sido muy cercanos a los teóricos. Estas predicciones son un indicador de que la selección de los atributos que describen cada elemento ha sido correcta. El *finger print* de cada elemento parece cumplir su función como se esperaba.

5. Electronegatividad de Pauling.

Llegados a este punto, contamos con una matriz de datos para cada elemento, once comunes y ocho independientes para cada estructura. Una de estas variables es la electronegatividad de Pauling, la cuál vamos a intentar predecir a través de los otros dieciocho restantes. Si hiciésemos una búsqueda de las correlaciones que presenta la electronegatividad de Pauling con el resto, obtendríamos que no existe ninguna variable con un índice de correlación alto.

Ante esta situación, nuestro análisis se realizará sobre subgrupos de la muestra considerada. La selección de dichos grupos la realizamos con diferentes criterios.

5.1. Predicción de la electronegatividad de Pauling y el radio d.

Vamos a realizar un agrupamiento de los elementos de acuerda al logaritmo en base 10 de la electronegatividad de Pauling y el radio d:

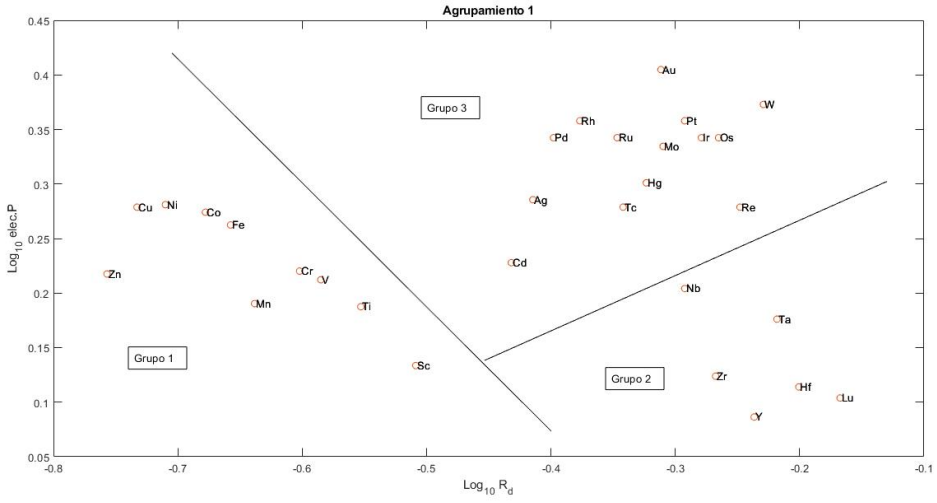

Figura 23: Agrupamiento según la electronegatividad de Pauling y el radio d en logaritmo de base 10.

- Grupo 1: Escandio, Titanio Vanadio, Cromo, Manganeso, Hierro, Cobalto, Níquel,

Cobre y Cinc.

- Grupo 2: Itrio, Circonio, Niobio, Lutecio, Hafnio y Tántalo.
- Grupo 3: Molibdeno, Tecnecio, Rutenio, Rodio, Paladio, Plata, Cadmio, Wolframio, Renio, Osmio, Iridio, Platino, Oro y Mercurio.

Recordemos que las dos variables son comunes para las tres estructuras, luego estos grupos son válidos para todas ellas. Si nos fijamos bien, podemos ver que no se trata de ningún agrupamiento aleatorio, ya que el grupo uno corresponde a los elementos de la primera fila, del Escandio al Cinc, el segundo grupo lo conforman los tres primeros elementos de la segunda y tercera fila, y el grupo tres lo forman los últimos siete elementos de las filas dos y tres. La idea de la obtención de la estimación es similar a la predicción de las curvas DOS. Obtener un conjunto de pesos mediante un ajuste de mínimos cuadrados. Esos pesos se combinan con los atributos de mayor correlación para obtener el valor estimado de la electronegatividad. La estimación se realiza sobre el orden de magnitud de las variables, por ello se pasará a trabajar sobre el logaritmo en base 10 de cada una de ellas. Esa elección se debe a obtención de un mejor resultado final.

El análisis de predicción lo realizaremos de la siguiente forma:

1. Se realizará independientemente para cada grupo de los tres recién obtenidos y, para cada grupo, habrá un análisis por cada estructura.
2. Primero obtenemos las covarianzas entre la electronegatividad de Pauling y el resto de variables, y de ellas fijaremos un valor mínimo, quedándonos con las variables con un índice de correlación superior a dicho mínimo. Como queremos que el sistema no sea indeterminado, el número de atributos involucrados ha de ser menor al número de elementos que intervienen en cada grupo.
3. Después pasamos a trabajar con los logaritmos en base diez de cada una de las variables seleccionadas, teniendo así una matriz de valores en la que cada fila representa

un elemento, y cada columna el valor del logaritmo de una variable. Denotaremos esta matriz por Xvar. Por ello solo podemos utilizar variables que sean estrictamente mayores que 0.

4. A continuación generamos la matriz F, que está compuesta por una primera columna de unos, de dimensión igual al número de elementos n, seguida de la matriz Xvar. La matriz F tendrá el mismo número de filas, más la primera columna de unos. Como esta vez no hemos centrado la matriz que representa las ecuaciones en el ajuste de mínimos cuadrados, introducimos el término independiente, es decir, el valor 1 con el que centramos la estimación. Al realizar esto de forma simultánea en todas las ecuaciones, aparece la columna de unos en la matriz F.

5. Resolvemos el sistema:

$$\mathbf{p} = F^{\dagger} \cdot \mathbf{b} \tag{39}$$

Donde $F^{\dagger}$ es la matriz pseudo inversa de Moore-Penrose de F:

$$F^{\dagger} = \left(F^t F\right)^{-1} F^t \tag{40}$$

Y **b** es un vector columna formado por el logaritmo en base 10 de la electronegatividad de Pauling. El vector **p** contiene los pesos de la predicción.

$$\mathbf{b} = \{b_1, ..., b_n\}^t \tag{41}$$

6. Por último, para obtener el valor de la predicción, resolvemos:

$$\mathbf{A} = F \cdot \mathbf{p} \qquad\qquad \mathbf{A} = \{a_1, ..., a_n\}^t \tag{42}$$

Hemos obtenido un vector con el logaritmo en base diez de la predicción. Deshaciendo el logaritmo:

$$E_i = 10^{a_i} \tag{43}$$

E_i Representa el valor de la aproximación para el elemento iésimo del grupo.

7. El error cometido en la predicción será:

$$\text{error} = \frac{\|\mathbf{b} - \mathbf{E}\|_\infty}{\|\mathbf{b}\|_\infty} \cdot 100 \tag{44}$$

Para los treinta elementos se hace un estudio de las cuatro estructuras, diferenciando cuando todos los elementos forman a estructura con la misma constante de red, denotado como misma constante de red, y cuando los elementos forman la estructura con la constante de red más estable para cada elemento, es decir, cada elemento tendrá una constante de red independiente. Para el grupo 1 se muestran los índices de correlación de cada variable con la electronegatividad de Pauling, primero con la misma constante de red, y después con la más estable. Los valores que aparecen en rojo no han sido utilizados en la predicción de la electronegatividad.

5.2. Análisis de los resultados.

El objetivo de esta sección es intentar determinar le electronegatividad, a partir del resto de atributos comunes para cada elemento. Debido a la poca correlación que existe entre la electronegatividad se ha optado por realizar un clasificación previa.

Comenzamos analizando el primer grupo obtenido al relacionar la electronegatividad con el radio d. Este grupo está compuesto por los elementos con número principal n=4.

La figura (24) compara los valores de a electronegatividad teórica con la obtenida en el grupo 1, para una estructura FCC obtenida con la constante de red más estable para cada elemento. Recordamos que los valores de la electronegatividad son adimensionales. La recta que aparece en la gráfica es la recta de pendiente 1, que representa los valores iguales entre la estimación y los valores teóricos. Como podemos ver, el ajuste es casi perfecto. El error

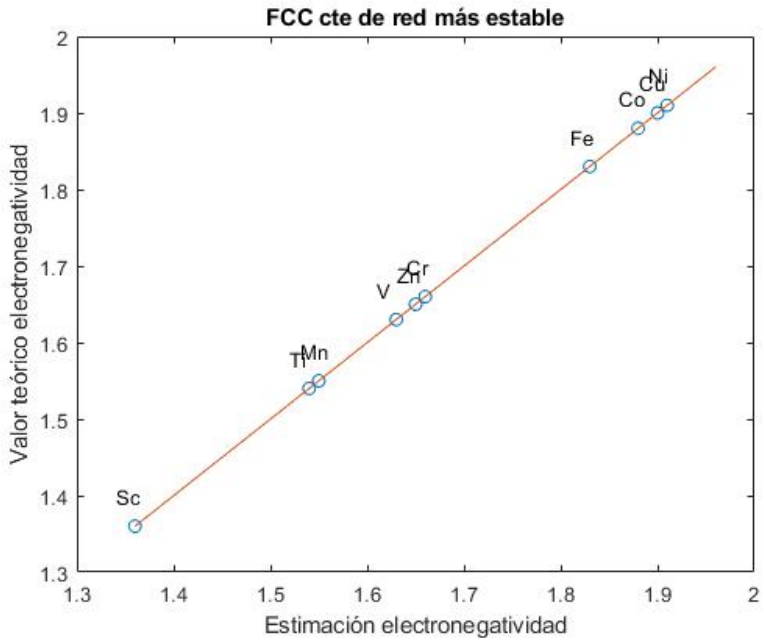

Figura 24: Comparación entre la electronegatividad teórica con la obtenida en el grupo 1 obtenido a través del radio del orbital d, para una estructura FCC compuesta por la constante de red más estable.

cometido en la estimación total es $8{,}23 \cdot 10^{-8}$. Como estamos trabajando con valores comprendidos entre 1.3 y 2.0, el error cometido es ínfimo.

Si analizamos las variables que están involucradas en este primer grupo considerando todas las estructuras, observamos que estas son comunes en prácticamente todas las estructuras, con un alto índice de correlación. La siguiente tabla recoge todos los valores del índice de correlación obtenidos. En rojo se muestran los valores que no han sido utilizados para la construcción de la estimación, pero al aparecer en otras estructuras pueden aportar información.

Grupo 1	Misma cte de red				Cte de red más estable			
Estructura	SC	FCC	BCC	HCP	SC	FCC	BCC	HCP
Variables								
Número atómico	0.7483	0.7483	0.7483	0.7483	0.7483	0.7483	0.7483	0.7483
E.Valencia	0.7483	0.7483	0.7483	0.7483	0.7483	0.7483	0.7483	0.7483
Radio S	-0.7353	-0.7353	-0.7353	-0.7353	-0.7353	-0.7353	-0.7353	-0.7353
Radio P	-0.7444	-0.7444	-0.7444	-0.7444	-0.7444	-0.7444	-0.7444	-0.7444
Radio D	-0.7943	-0.7943	-0.7943	-0.7943	-0.7943	-0.7943	-0.7943	-0.7943
Mendeleev Number	0.7788	0.7788	0.7788	-0.7943	0.7788	0.7788	0.7788	0.7788
Escala Química	-0.7438	-0.8028	-0.7941	-0.7943	-0.7438	-0.8028	-0.7941	-0.8028
Anchura	-0.7794	-0.7619	-0.8060	-0.7943	-0.3933	-0.6521	-0.6889	-0.7619
Pts Críticos	-0.7531	-0.7998	-0.8283	-0.7943	-0.2675	-0.6860	-0.7313	-0.7998
Error (%)	2.42E-07	1.09E-07	1.34E-08	1.09E-08	3.7491	8.23E-08	9.40E-08	1.09E-07

Cuadro 3: Valores del coeficiente de correlación con la electronegatividad del Grupo 1 obtenido a través del radio del orbital d.

La predicción en este grupo ha sido excelente. Independientemente de la estructura el error ha sido mínimo, y el hecho de que la estructura no influya en la predicción, le da un carácter más general.

Pasamos ahora al análisis del grupo 2. Este grupo tiene el problema de que el número de elementos involucrado es mucho menor. Como nos interesa que al resolver el sistema de ecuaciones que involucra la matriz F, sea determinado, ya que de ser indeterminado, la estimación podría no ser válida, el número de variables no puede ser superior al número de elementos. Esto provoca que aunque haya variables con alto valor de correlación, no se consideren en la estimación. Esto tiene como consecuencia una discrepancia entre las variables involucradas en cada estructura. Aunque la estimación es muy buena, incluso superior que en el grupo 1, su falta de un carácter general la hace menos atractiva. Podemos destacar la fuerte relación que existe con los electrones de valencia. El uso de atributos específicos de cada estructura con un índice de correlación muy alto, desplaza a otros atributos comunes, como por ejemplo el radio p. Aunque su valor es muy alto, la limitación del número de atributos que se utilizan en la reconstrucción hace que hay un gran número de variables, y

pocas de ellas comunes para todas las estructuras.

Grupo 2	Misma cte de red				Cte de red más estable			
Estructura	SC	FCC	BCC	HCP	SC	FCC	BCC	HCP
Variables								
E.Valencia	0.9330	0.9330	0.9330	0.9330	0.9330	0.9330	0.9330	0.9330
Radio S	-0.8758	-0.8758	-0.8758	-0.8758	-0.8758	-0.8758	-0.8758	-0.8758
Radio P	-0.8514	-0.8514	-0.8514	-0.8514	-0.8514	-0.8514	-0.8514	-0.8514
Escala Química	0.8707	0.8707	0.8707	0.8707	0.8707	0.8707	0.8707	0.8707
Cte de Red	-0.8934	-0.8951	-0.8932	-0.8951	-0.8934	-0.8951	-0.8932	-0.8951
P.Inicial DOS	-0.0376	-0.8727	-0.7378	-0.7473	-0.7234	-0.8009	-0.7384	-0.6667
P.Final DOS	-0.7571	-0.9196	-0.9225	-0.9231	-0.7678	-0.8372	-0.7560	-0.8112
Anchura	-0.8324	-0.8245	-0.9071	-0.8583	-0.6406	-0.5006	-0.6477	-0.4046
Max. DOS	0.9539	0.5840	0.7879	0.9373	-0.7996	-0.2231	-0.8819	-0.6258
E. Max DOS	-0.5702	-0.0085	-0.9337	0.2254	-0.9287	-0.3939	-0.2411	-0.3162
Dif Max min DOS	-0.4148	0.2301	0.6724	0.8668	-0.7910	-0.6589	-0.9114	-0.6240
Pts Críticos	-0.9436	-0.8209	-0.6964	-0.7931	0.2252	-0.6619	-0.5223	-0.7862
Error (%)	1.26E-10	4.53E-11	4.00E-09	1.77E-10	2.75E-10	2.65E-10	6.53E-10	2.65E-10

Cuadro 4: Valores del coeficiente de correlación con la electronegatividad del Grupo 2 obtenido a través del radio del orbital d.

El grupo 3 es el que presenta el mayor error en la estimación. Los valores de la correlación son muy bajos, y esto ha provocado una mala predicción de la electronegatividad.

Grupo 3		Misma cte de red				Cte de red más estable		
Estructura	SC	FCC	BCC	HCP	SC	FCC	BCC	HCP
Variables								
Radio S	0.6492	0.6492	0.6492	0.6492	0.6492	0.6492	0.6492	0.6492
Radio P	0.5597	0.5597	0.5597	0.5597	0.5597	0.5597	0.5597	0.5597
P.Inicial DOS	0.1466	-0.5115	-0.6040	-0.5464	-0.5623	-0.0096	-0.5738	-0.4638
P.Final DOS	0.3704	0.4467	0.5046	0.5022	0.6020	0.4646	0.4842	0.4636
Max. DOS	0.4367	0.4610	0.4666	0.4516	0.5902	0.4863	0.5135	0.4572
E. Max DOS	0.3365	-0.4144	-0.5868	-0.4928	-0.5684	-0.4549	-0.5550	-0.5033
Dif Max min DOS	0.4075	0.0161	-0.4849	-0.2146	-0.5288	-0.5881	-0.4957	-0.5034
Pts Críticos	0.4424	0.4512	0.5379	0.4573	0.4739	0.4329	0.5176	0.4255
Error (%)	11.9525	11.4607	8.7587	11.9322	9.5282	12.4023	9.2002	11.8099

Cuadro 5: Valores del coeficiente de correlación con la electronegatividad del Grupo 3 obtenido a través del radio del orbital d.

El buen comportamiento del grupo 1, a diferencia de los otros dos ha motivado a realizar un análisis por filas. El grupo 1 se mantiene, mientras que el grupo 4 ahora lo componen los elementos con número principal n=5, y el grupo 5 los elementos de n=6.

Los valores para el grupo 4 son:

Grupo 4	Misma cte de red				Cte de red más estable			
Estructura	SC	FCC	BCC	HCP	SC	FCC	BCC	HCP
Variables								
Número atómico	0.6191	0.6191	0.6191	0.6191	0.6191	0.6191	0.6191	0.6191
E.Valencia	0.6191	0.6191	0.6191	0.6191	0.6191	0.6191	0.6191	0.6191
Radio S	-0.5743	-0.5743	-0.5743	-0.5743	-0.5743	-0.5743	-0.5743	-0.5743
Radio P	-0.6878	-0.6878	-0.6878	-0.6878	-0.6878	-0.6878	-0.6878	-0.6878
Escala Química	0.7230	0.7230	0.7230	0.7230	0.7230	0.7230	0.7230	0.7230
P.Inicial DOS	0.5512	0.5512	0.5512	0.5512	0.5512	0.5512	0.5512	0.5512
P.Final DOS	-0.8990	-0.9281	-0.9144	-0.9281	-0.8990	-0.9281	-0.9144	-0.9281
Anchura	-0.6396	-0.6619	-0.6479	-0.6566	-0.3677	-0.3878	-0.4754	-0.3849
Max. DOS	0.6947	0.4711	0.2889	0.4048	-0.0709	0.1190	-0.2241	-0.0169
E. Max DOS	-0.6433	-0.1083	-0.4786	0.2456	-0.5721	-0.3179	-0.3406	-0.0731
Dif Max min DOS	0.6686	0.6301	0.3501	0.5473	0.1566	0.0575	0.0927	0.1045
Pts Críticos	-0.6412	-0.7805	-0.7612	-0.7436	-0.3725	-0.5292	-0.6059	-0.5629
Error (%)	7.28	1.75E-07	0.3907	5.72E-07	0.8059	0.8482	0.2094	0.7642

Cuadro 6: Valores del coeficiente de correlación con la electronegatividad del Grupo 4 obtenido al considerar los elementos con número principal n=5.

Aunque su error es mayor que el calculado en el grupo 2 anterior, sigue siendo lo suficientemente bajo como para que el resultado sea aceptable. Esta vez las variables involucradas no presentan tanta discrepancia. Si nos fijamos en la tabla (6), los atributos que no son comunes son introducidos en su mayoría por la estructura SC formada con la misma constante de red para cada elemento, quien es precisamente la de mayor error. Concluimos que la predicción en dicha SC no ha sido tan buena como las otras, aún teniendo un error pequeño, y las 7 estructuras restantes tienen muy buen comportamiento.

Por último analizamos el grupo 5.

Grupo 5	Misma cte de red				Cte de red más estable			
Estructura	SC	FCC	BCC	HCP	SC	FCC	BCC	HCP
Variables								
Número atómico	0.7863	0.7863	0.7863	0.7863	0.7863	0.7863	0.7863	0.7863
E.Valencia	0.7863	0.7863	0.7863	0.7863	0.7863	0.7863	0.7863	0.7863
Radio P	-0.7458	-0.7458	-0.7458	-0.7458	-0.7458	-0.7458	-0.7458	-0.7458
Radio D	-0.8136	-0.8136	-0.8136	-0.8136	-0.8136	-0.8136	-0.8136	-0.8136
Mendeleev Number	0.776	0.776	0.776	0.776	0.776	0.776	0.776	0.776
Escala Química	0.7305	0.7305	0.7305	0.7305	0.7305	0.7305	0.7305	0.7305
Cte de Red	-0.6066	-0.6488	-0.6171	-0.6488	-0.6066	-0.6488	-0.6171	-0.6488
P.Inicial DOS	-0.2678	-0.5771	-0.659	-0.5945	-0.8179	-0.8779	-0.8868	-0.5771
P.Final DOS	-0.8241	-0.7158	-0.7221	-0.7013	-0.6622	-0.6895	-0.7101	-0.7158
Anchura	-0.8544	-0.7636	-0.7649	-0.7414	-0.4153	-0.4855	-0.4261	-0.7636
E. Max DOS	-0.3915	-0.6535	-0.7298	-0.2376	-0.87	-0.7274	-0.6783	-0.6535
min. DOS	0.8644	0.8288	0.7	0.5364	0.3471	0.1496	0.1879	0.8288
Dif Max min DOS	0.0728	0.561	0.7228	0.7474	-0.0521	0.3079	-0.0528	0.5610
Pts Críticos	-0.8466	-0.8363	-0.761	-0.769	-0.4638	-0.5045	-0.4089	-0.8363
Error (%)	3.96E-07	3.92E-06	6.98E-05	8.12E-06	4.67E-06	1.29E-05	1.62E-05	1.04E-04

Cuadro 7: Valores del coeficiente de correlación con la electronegatividad del Grupo 5 obtenido al considerar los elementos con número principal n=6.

Los errores de la predicción han sido muy bajos. No se ha logrado conseguir el carácter general independiente de la estructura buscado. Sin embargo, dentro de las cuatro estructuras formadas por la constante de red más estable, las variables sí que aparecen con valores similares.

En general se concluye que en la predicción de la electronegatividad, se han conseguido buenos resultados. Los errores han sido pequeños, y aunque no siempre ha podido conseguirse un comportamiento de las variables independiente de la estructura, la selección por filas de los grupos 1, 4 y 5 ha obtenido un mejor resultado que los grupos 1, 2 y 3.

6. Clasificación de los elementos.

El siguiente objetivo es realizar una clasificación de los elementos. Recordamos que es de esperar que elementos similares produzcan materiales con características similares. La clasificación se ha realizado teniendo en cuenta los elementos con las cuatro estructuras, con la misma constante de red y diferente constante de red, y se han añadido dos conjuntos de datos más. El primero contiene los elementos con la estructura en la que se encuentran en la naturaleza. Cada elemento tendrá una constante de red diferente, en concreto la más estable, y en el conjunto total aparecen elementos con estructura SC, BCC, FCC y HCP.

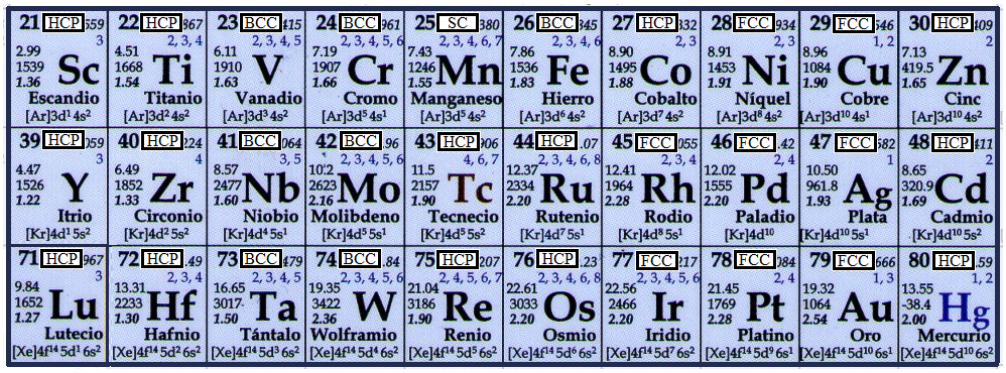

Figura 25: Estructura más común de cada elemento que presenta en la naturaleza.

La clasificación que se obtiene de este conjunto tiene especial importancia, dado que es la más coherente desde un punto de vista físico.

El segundo conjunto que se ha añadido corresponde a todas las estructuras con la misma constante de red y diferente constante de red a la vez. Es decir, realizamos un estudio sobre todos los elementos a la vez, en el que cada elemento tiene ocho clones. El tratamiento de esta clasificación es un poco especial. La matriz que contiene las dos primeras coordenadas PCA, es $CPCA \in M_{30xq}(\mathbb{R})$, sin embargo esta matriz se había calculado para una estructura concreta de 30 elementos. Por lo que ahora nuestra matriz $CPCA \in M_{240xq}(\mathbb{R})$. Nuestro objetivo es optimizar el número de clones de un elemento que pertenece a un grupo concreto. Para ello buscamos el centro de gravedad que genera los clones de cada elemento. Dividimos la matriz $CPCA$ en 30 submatrices de dimensión 8 x q, en la que cada submatriz representa a cada elemento por separada. Cada submatriz de $CPCA$ la denotemos como S_i, donde i

toma valores entre 1 y 30. Las filas de S_i representan a los clones, mientras que las columnas las coordenadas PCA. Para hallar el centro de gravedad de las coordenadas PCA, realizamos la media de cada columna. Por ejemplo para el Escandio, tenemos:

$$S_{Sc} = \frac{1}{8}\sum_{j=1}^{q}\sum_{i=1}^{8} s_{i,j} \ \in M_{1\times q}(\mathrm{R}), \tag{45}$$

donde $s_{i,j}$ es el elemento de la fila i, columna j de S.

Ahora que tenemos 30vectores que representan los centros de gravedad por coordenadas PCA de cada elemento por separado, podemos volver a formar una matriz $CG \in M_{30xq}$, en la que la fila iésima de CG, es S_i. Por último ya solo queda clasificar los elementos por grupos de la misma forma que se ha hecho con las diferentes estructuras.

En el punto en que nos encontramos ahora, tenemos 10 clasificaciones diferentes de los elementos en cinco grupos diferentes. El número de grupos ha sido obtenido gracias al algoritmo de clusterización. El siguiente paso es determinar una única clasificación final. Para ello nos fijamos que hay muchos elemento que siempre tienden a juntarse. Por ejemplo, el Escandio, el Titanio y el Vanadio siempre aparecen en el mismo grupo. El Zinc el Cadmio y el Mercurio suelen agruparse juntos, aunque en algunas estructuras alguno de ellos formaba un grupo individual, por lo que, como interesa trabajar con grupos de más de un elemento, unimos estos dos grupos de forma que en la clasificación total pasamos de cinco grupos a cuatro.

Una vez que se han caracterizado cuales son los elementos más representativos de cada grupo, escogemos a que grupo pertenece cada elemento restante maximizando el número de veces en el que ese elemento pertenece a un grupo concreto entre las diez clasificaciones.

Los elementos más representativos de cada grupo son. El Escandio para el grupo 1, el Hierro para el grupo 2, el Oro en el grupo 3 y el Mercurio en el grupo 4. Estos elementos pertenecen a grupos diferentes en prácticamente todas las estructuras. Ahora si nos fijamos en un elemento en concreto, por ejemplo en la Plata, vemos que pertenece al grupo de Oro, a excepción de la estructura SC con la misma constante de Red para cada elemento. La

plata pertenecerá al grupo del Oro, el grupo 3. Realizando este proceso para cada elemento tenemos:

- Grupo 1: Escandio, Titanio, Vanadio, Cromo, Itrio, Circonio, Niobio, Molibdeno, Lutecio, Hafnio y Tántalo.
- Grupo 2: Manganeso, Hierro, Cobalto, Níquel, Cobre, Tecnecio, Rutenio, Rodio, Paladio, Wolframio, Renio, Osmio e Iridio,
- Grupo 3: Plata, Platino y Oro.
- Grupo 4: Cinc, Cadmio y Mercurio.

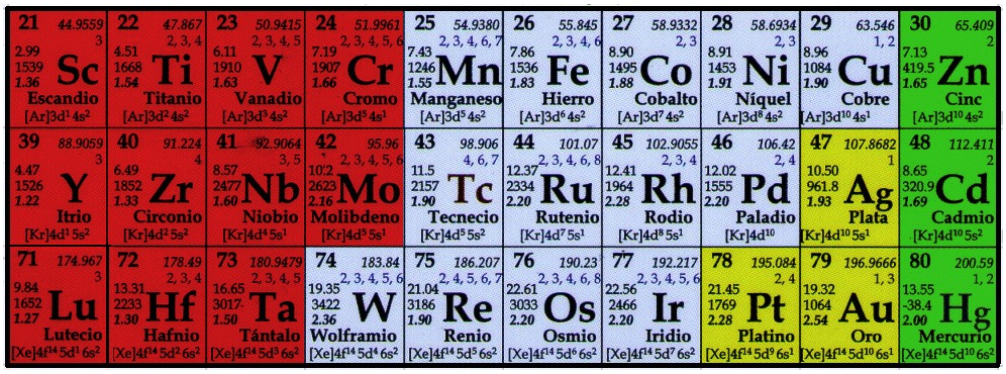

Figura 26: Clasificación final de los elementos.

Ahora a partir de los atributos que no dependen de las estructuras, caracterizamos cada grupo con las variables que tengan sentido. Por ejemplo, el número atómico crece desde 21 hasta 80, y dado que elementos del mismo grupo, tienen una variación del número atómico muy grande, este parece no ser un factor decisivo en la clasificación. La figura () indica el valor de la media μ y la desviación típica σ de cada grupo.

Variables		E. Valencia	Elec. Pauling	R_s (Å)	R_p (Å)	R_d (Å)	M.N.	E.Q.
Grupo 1	μ_1	4.36	1.51	1.22	1.52	0.47	43.73	0.78
	σ_1	1.12	0.27	0.09	0.10	0.16	15.41	0.08
Grupo 2	μ_2	8.38	2.02	1.07	1.35	0.38	63.23	1.02
	σ_2	1.45	0.23	0.12	0.15	0.15	4.57	0.09
Grupo 3	μ_3	10.67	2.25	1.17	1.41	0.46	69.67	1.15
	σ_3	0.58	0.31	0.11	0.07	0.07	1.53	0.04
Grupo 4	μ_4	12.00	1.68	0.93	1.17	0.31	75.33	1.39
	σ_4	0.00	0.02	0.10	0.10	0.11	0.58	0.05

Cuadro 8: Tabla de valores resumen de los grupos.

6.1. Análisis de los resultados.

Tras realizar las diferentes clasificaciones, hemos obtenido como resultado una clasificación final de los elementos de estudio. Hemos obtenido cuatro grupos, el grupo 1, de 11 elementos. El grupo 2 de 13 elementos. Y los grupos 3 y 4 de 3 elementos cada uno. La clasificación final es muy similar a la obtenida usando las curvas de la estructura que aparecía en la naturaleza de cada elemento. El hecho de que esta clasificación, que era la que más información podía dar, sea muy parecida a la final, es un buen indicador de haber obtenido una clasificación correcta.

El uso de las diferentes estructuras y diferentes constantes de red, utilizadas de forma independiente, o conjuntamente, ha provocado una mayor robustez en el resultado final. De esta forma, se ha conseguido un resultado mucho más universal que es capaz de considerar diferentes situaciones.

Si nos centramos en identificar qué atributos hacen que cada elemento esté en ese grupo en concreto, podemos observar gracias a la tabla (8), que los electrones de valencia son un buen indicador. Los valores medios de cada grupo son bastante diferentes, y su desviación típica no es muy grande. Los grupos 3 y 4 son los que mejor cumplen esto. En el grupo 1

los elementos tienen su última subcapa, el orbital d, un llenado que va desde 1 hasta 5. En el grupo 2 este llenado va desde 5 pero no consigue completar la subcapa con 10 electrones. En este sentido todo parece indicar que los elementos del grupo 1 tienen un llenado hasta la mitad, mientras que el grupo dos es un llenado superior a la mitad. La diferencia entre el grupo 3 y el 4 es que el 4 es capaz de llenar todas sus subcapas, la d y la s, mientra que al 3 le falta uno o dos electrones.

Los valores medios de la electronegatividad son diferentes, con una desviación no muy grande. Sin embargo, no parece que siga un orden concreto para cada grupo.

El comportamiento del radio atómico en los orbitales s,p y d, nos indica que el radio decae a medida que avanzamos de izquierda a derecha o de abajo a arriba. Como nuestros grupos consideren elementos de la misma fila y de la misma columna, aunque los radios sean similares, y la desviación pequeña, no es un buen atributo que caracterice cada grupo.

El número de Mendeleev aumenta en cada grupo, a pesar de tener una desviación grande en el primer grupo. Puede ser un buen atributo característico del grupo, ya que en los grupos en los que es más parecido, los grupos 2, 3 y 4, su pequeña desviación podría ayudar a clasificar un nuevo elemento.

Por último, la escala química si parece ser un buen indicador de grupo. Su valor es creciente en cada grupo, y su desviación es pequeña. La relación entre la escala química y los orbitales más externos, hace que esta variable se relacione con los electrones de las subcapas exterior. Este hecho concuerda con los electrones de valencia quienes también eran buenos caracterizadores.

La conclusión final que intenta explicar el la formación de estos grupos, es que el número de los electrones en las capas exteriores y su posición, son de gran importancia. Esto justificaría por qué todas las estructuras daban clasificaciones parecidas, y destaca el hecho que las clasificaciones, a pesar de haberse realizado con diferentes estructuras, no depende de la

forma en la que los elementos se disponen en el sólido final.

7. Predicción del Circonio.

El objetivo de este apartado es utilizar todos los algoritmos y resultados desarrollados hasta ahora. Para ello, eliminamos de nuestra base de datos un elemento, por ejemplo, el Circonio. Nuestro objetivo ahora, consiste en elaborar un perfil de atributos del Circonio, es decir, su *finger print*, predecir su densidad de estados en una estructura concreta y predecir su electronegatividad en la escala de Pauling. La nueva base de datos está formada por los 29 elementos restantes. El uso del Circonio, y no de otro elemento externo, permite comparar los resultados obtenidos con los teóricamente correctos que ya conocemos.

El primer paso en el estudio del Circonio, parte de los atributos que se han obtenido de la bibliografía. Estos son:

	Zirconio	Grupo 1	
Variables		μ	σ
Número atómico	40	-	-
E.Valencia	4	4.36	1.12
Quantum Number	5	-	-
Radio S (Å)	1.265	1.220	0.09
Radio P (Å)	1.560	1.520	0.1
Radio D (Å)	0.540	0.470	0.16
Mendeleev Number	49	43.73	15.41
Escala Química	0.76	0.78	0.08

Cuadro 9: Atributos conocidos del Circonio. A la derecha se muestran los valores que caracterizan el Grupo 1.

Ahora, gracias a la clasificación de la tabla (8), identificamos a que grupo corresponde el Circonio. Todos los valores encajan perfectamente con el Grupo 1, incluidos en la tabla para

facilitar la selección del grupo.

Una vez identificado el grupo, pasamos a trabajar con los elementos ya conocidos que pertenecen a dicho grupo (elementos de la figura (26) a excepción del Circonio). Identificamos este grupo como Grupo Zr.

Como conocemos las curvas DOS de los elementos del Grupo Zr, podemos generar las matrices $CPCA$ y $PCAR$ necesarias para la predicción del la curva DOS [4,1]. De la misma forma podemos obtener el vector media μ generado en la ecuación (25). Tanto las dos matrices como la media se obtienen al considerar los diez elementos del Grupo Zr, considerando una estructura en concreto. En este caso, se utilizará la estructura FCC generada por la constante de red más estable para cada elemento como ejemplo, pero el proceso es análogo para el resto. La idea de la predicción es sencilla. Generar unos pesos gracias a los elementos de los que conocemos las curvas y a través de ellos, reconstruir una curva DOS con los atributos del Circonio.

Para generar la predicción realizamos el mismo proceso mostrado en la sección 4, con predicciones simultáneas que variaban la selección de elementos y de atributos. En este caso el número mínimo de elementos es igual a 8. Generamos los pesos como se mostró en la ecuación (35) utilizando los atributos de los elementos seleccionados. Dichos atributos son los descritos en el capítulo 2, que han sido obtenidos de la bibliografía. Los pesos generados han utilizado la matriz $CPCA$, que acabamos de obtener en este apartado.

Introducimos ahora una modificación con el algoritmo de la sección 4. La matriz de atributos que vamos a utilizar ahora no es la misma utilizada a la hora de generar los pesos. En este caso, vamos a utilizar una matriz de atributos en la que todas las filas son iguales y corresponden a los atributos del Circonio. Con esto conseguimos, una vez conseguidos los pesos, y la matriz $PCAR$ obtenida anteriormente, la predicción de la curva del Circonio. Realizamos la media de las predicciones obtenidas como se explicó en el algoritmo y obtenemos la curva DOS:

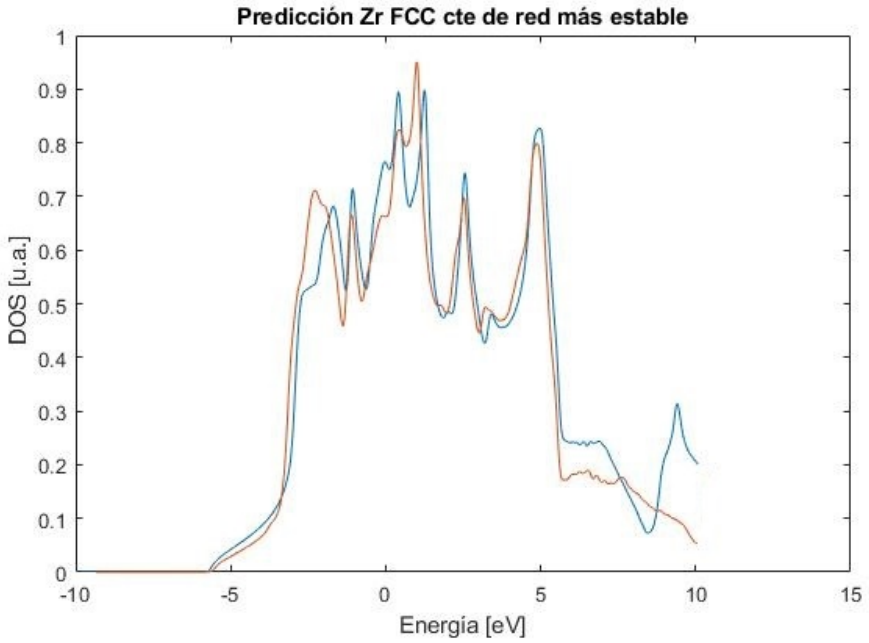

Figura 27: Predicción del Circonio en estructura FCC generado con la constante de red más estable para cada elemento. En azul la curva teórica, en rojo la aproximación.

La figura (27) muestra la predicción de la curva DOS del Circonio, en rojo, y la curva teórica DOS, en azul.

Una vez obtenida la curva, realizamos el mismo análisis de atributos presentada en la sección [4.2], donde se incluye la anchura de la curva, el valor del máximo y el mínimo etc. Con esto completamos el *finger print* del Circonio, a excepción de la electronegatividad, la cual vamos a predecir ahora.

	P.Inicial DOS (eV)	P.Final DOS (eV)	Anchura (eV)	MAX. DOS	E. Max DOS (eV)	min. DOS	Dif Max min DOS	Pts Cr ıticos
Predicción	-2.7375	5.1914	7.929	0.9424	1.0143	0.4519	0.4905	25
Teóricos	-2.7873	5.4168	8.2041	0.8981	1.2648	0.4261	0.4720	17
Error(%)	1.7867	4.1611	3.3532	4.9326	19.8055	6.0549	3.9195	47

Cuadro 10: Comparación de los atributos obtenidos entre la DOS teórica y la predicción.

La idea de la predicción de la electronegatividad es muy similar a la predicción de la curva. El algoritmo de predicción de la electronegatividad dentro de un grupo, generaba unos

pesos (ecuación (39)), que luego se utilizaban para generar la predicción, combinándolos con los atributos. Los atributos que intervienen en el Grupo Zr en esta predicción son:

Variables	Coeficiente de correlación
E.Valencia	0.8656
Radio S	-0.5713
Radio P	-0.6022
Número de Mendeleev	0.7099
Escala Química	0.832
Cte de Red	-0.777
P.Inicial DOS	-0.5567
P.Final DOS	-0.7989
min. DOS	0.6259

Cuadro 11: Variables y su coeficiente de correlación utilizadas en la predicción de Pauling.

Para generar los nuevos pesos, realizamos el algoritmo de predicción de electronegatividad sobre el Grupo Zr, obteniendo así los pesos, $\mathbf{p}$.

Recordamos que la predicción se realizaba sobre el logaritmo en base 10 de los atributos, así que generamos el vector de atributos del Circonio en logaritmo en base 10, $\mathbf{Zr}$. Este vector contiene el logaritmo en base 10, del $finger\ print$ conocido hasta ahora del Circonio, seleccionando solo las variables de la tabla (11).

Para obtener la predicción, operamos:

$$\text{Pauling Zr} = 10^{\mathbf{Zr}\cdot\mathbf{p}} \tag{46}$$

El valor obtenido de la predicción es 1.3877. El valor teórico es de 1.3300, con lo que se ha cometido un error del 4.33 %.

7.1. Análisis de los resultados.

En esta sección se ha puesto en común todos los resultados generados en este trabajo. En primer lugar, la predicción de la curva DOS ha sido acertada. El cuadro (10) muestra que los atributos obtenidos de la curva predicha, son muy similares a los obtenidos en la curva teórica. Exceptuando los puntos críticos, la figura (27) muestra un buen ajuste, sobretodo en la parte final de la zona de mayor interés, la zona central.

El buen comportamiento de esta predicción indica que la selección de los elementos para formar los grupos y elaborar la clasificación final, ha sido correcta. Haber obtenido una curva DOS similar a la teórica, se debe a utilizar elementos con propiedades similares. La idea de que materiales similares tendrían propiedades similares era uno de los principales objetivos del trabajo, y se concluye que efectivamente, en esta investigación, se cumple.

La predicción de la electronegatividad parte de las predicciones anteriores. El resultado final ha sido muy satisfactorio. Con tan solo un error del 4.33 %, se ha logrado predecir su valor, utilizando datos a su vez predichos anteriormente.

8. Conclusiones.

El objetivo de este trabajo es estudiar las propiedades de los materiales en función de las variables que describen los elementos que los componen. Para lograrlo, se planteó el problema de elaborar un perfil de atributos para cada elemento, denominado *finger print*, y predecir su curva DOS, la cual juega un papel fundamental en las propiedades de los sólidos. En el desarrollo de esta investigación, se tiene en cuenta que elementos con características similares, producen materiales similares.

La base de datos está compuesta por la curvas DOS de 30 elementos químicos correspondientes a los metales de transición comprendidos entre el Escandio y el Mercurio, considerando 4 estructuras en las que se presentan en la naturaleza, SC, BCC, FCC y HCP. Cada estructura puede estar formada o bien por por la constante de red más estable para cada elemento, o por la misma constante de red, fija para todos los elementos. La constante de red es el valor de la arista que compone el prisma que mediante traslaciones genera toda la estructura. De esta forma la base de datos la forman 8 conjuntos de datos, caracterizados por la estructura.

El primer paso de este análisis fue elaborar el perfil de los elementos. El primer grupo de atributos que componen el perfil de cada elemento está formado por variables que ya han tenido un fuerte análisis en la física. Estas son: el número atómico, el cual indica el número de protones en un átomo. El número de electrones de valencia, que indica el número de electrones situados en las capas no completas del átomo. Número cuántico de la capa de valencia, que proporciona el nivel energético dado por el mayor valor del número principal n en la configuración. La electronegatividad de Pauling, que se define como la capacidad de un átomo de atraer electrones hacia si mismo. Los radios orbitales s, p y d los cuales dan un valor medio del tamaño del orbital correspondiente. La escala química, que nos da información sobre las estructuras cristalinas en compuestos binarios. La contante de red. Y por último la energía de Fermi que nos da la diferencia de energías entre el estado más alto y más bajo de un sistema cuántico a 0 K.

Con el objetivo de eliminar la subjetividad en el agrupamiento de elementos y conseguir un mayor rigor en el análisis, se desarrolló un algoritmo de clusterización que identifica el número de grupos óptimo y los elementos que pertenecen a dicho grupo. Este algoritmo tiene como punto de partida el algoritmo k-means ++, que agrupa los elementos minimizando la distancia de aquellos que pertenecen al mismo grupo. Mediante una selección de un centro inicial escogido de manera aleatoria, se forma el conjunto de centros que mediante repetidas iteraciones llega a un conjunto de centros finales, a los que se les asignan los elementos por cercanía. Evaluando simultáneamente agrupamientos con un número diferente de grupos, se busca la clasificación que maximiza la distancia entre grupos, minimizando la distancia de los elementos que componen cada uno de los grupos. Debido al factor aleatorio del algoritmo base, el resultado del número de grupos óptimo es variable, pero en un rango de valores pequeño que sí aporta mucha información. Como solución a este problema se llegó a la conclusión de, o bien escoger el grupo subjetivamente de acuerdo a criterios propios del problema, o en cambio, realizar el algoritmo un número grande de veces y escoger la moda de los resultados.

El análisis de la curva DOS pasó por la realización de una base PCA. Con esto se logró minimizar la dimensión de la muestra maximizando la información. Las nuevas variables obtenidas no correlacionadas linealmente permitieron una primera clasificación de acuerdo a las dos primeras variables de la base PCA. Estas son las que presentan una máxima variación y por tanto, son más adecuadas a la hora de realizar un agrupamiento. Gracias a esto, y con ayuda de un algoritmo de análisis de atributos sobre las curvas DOS, se obtuvieron el resto de atributos que completan el perfil de los elementos.

En cada curva se seleccionó su sección central, que aporta mayor información. En esta región se obtuvieron las energías inicial y final, y su diferencia, dando así el ancho de la curva de interés. Se obtuvo el valor máximo de la curva DOS, y su energía. También se obtuvo el valor mínimo, y se hayo la diferencia entre el máximo y el mínimo de la curva DOS. Por último, se midió el número de puntos críticos que aparecían en la curva de interés.

En este punto se completó la obtención de atributos, 11 comunes a todas las estructuras y obtenidos por bibliografía, y 8 independientes de la estructura obtenidos del análisis de las DOS. Gracias a esto, y mediante la información obtenida de la base PCA, se desarrolló un algoritmo de regresión en el cual se generaban unos pesos que combinados con el perfil de atributos se obtenía una predicción de la curva DOS. Este algoritmo considera predicciones simultáneas en las que varían los elementos y atributos involucrados en la generación de los pesos. El valor promedio de las predicciones simultáneas tuvo un comportamiento adecuado. El número mínimo de elementos involucrados en cada predicción juega un papel importante, consiguiendo un buen resultado, en todas las estructuras, utilizando al menos la mitad de elementos involucrados en la predicción de la muestra que se quiere predecir.

El siguiente paso fue, predecir el valor de uno de los atributos que componen el perfil, la electronegatividad de Pauling. La falta de correlaciones entre esta variable y el resto, motivó a un análisis por grupos obtenidos al comparar el radio d con la electronegatividad. Se obtuvieron tres grupos y mediante un algoritmo de regresión, a través del logaritmo de los atributos de los elementos, se obtuvo una estimación de la electronegatividad. En general los resultados obtenidos fueron muy satisfactorios. El excelente comportamiento del grupo 1, compuesto por los metales de la final 4 de la tabla periódica, motivó un análisis con grupos diferentes. Estos nuevos tres grupos estaban compuestos por las tres filas de la base de datos, de acuerdo a la tabla periódica. Los resultados fueron aún mejores. El error cometido fue despreciable y, en algunos casos, se obtuvo que los atributos involucrados en la predicción eran comunes a todas las estructuras, consiguiendo un carácter global de gran interés.

Por último se estableció una clasificación de los elementos final. Esta clasificación se realizó a través de las coordenadas de la base PCA obtenidas de las 8 estructuras a las que se añadieron dos conjuntos de datos más. El primero estaba compuesto por los elementos en la estructura en la que aparecen en la naturaleza, es decir, se consideraron los 30 elementos pero esta vez, aparecían estructuras SC, BCC, etc formadas por la constante de red más estable en cada elemento. El segundo conjunto de datos fue una combinación de las

8 estructuras a la vez, en el que se tenía 8 clones de cada elemento. Después se realizó la base PCA y mediante el centro de gravedad de cada elemento (con sus 8 clones), se obtuvo una nueva clasificación. De la combinación de las 10 clasificaciones anteriores, se obtuvo una clasificación final.

El último apartado de este trabajo junta todo lo desarrollado anteriormente. Se consideró al Circonio como un elemento en el que sólo se conocía de él los atributos fácilmente obtenibles de la bibliografía, a excepción de la electronegatividad de Pauling. Primero se identificó el grupo al que pertenecía de los obtenidos en la clasificación final. A continuación, utilizando solo los elementos de dicho grupo, y gracias al algoritmo de predicción de la curva DOS, se construyó un conjunto de pesos, que combinados con los atributos conocidos del Circonio, lograron obtener la predicción de la curva DOS del Circonio en la estructura FCC formada por la constante de red más estable para cada elemento. En este punto se realizó la obtención de atributos a partir de la curva DOS, completando el perfil del Circonio. Por último gracias al algoritmo de predicción de la electronegatividad, se generó un conjunto de pesos utilizando solo los elementos del grupo del Circonio. Combinando estos pesos con el perfil del Circonio, se obtuvo una estimación de la electronegatividad, con tan solo un error del 4.33 %. El conjunto de resultados de este apartado que desembocaron en un éxito final, demuestran el buen comportamiento de las herramientas desarrolladas en el trabajo. Tanto el trabajo de clasificación como los algoritmos de predicción han cumplido su papel.

En este trabajo se han realizado un análisis mediante técnicas matemáticas y de inteligencia artificial sobre una base de datos pertenecientes a la física de la materia condensada. Se ha optado por comprender la naturaleza de los datos con los que se iba a trabajar y de utilizar algoritmos y herramientas que, aunque no fuesen muy complejos, obtuviesen unos resultados coherentes e interpretables. Se ha dado más importancia al estudio de las herramientas y su comprensión que al uso de técnicas más avanzadas que, pudiendo obtener resultados aún mejores, se tuviese sobre ellas un menor control.

Las matemáticas son una herramienta muy poderosa que, una vez más, ha demostrado

un comportamiento extraordinario en otro ámbito, como es la física. En el transcurro de este trabajo han surgido nuevos problemas no contemplados inicialmente, que han sido resueltos de forma satisfactoria. Esta investigación se inició en un campo del que se encontró poca bibliografía, en concreto, en cuanto a la clasificación de los elementos gracias a la curva DOS, así como su predicción.

Los objetivos de este trabajo se cumplieron y en este punto sólo queda pensar en posibles nuevas vías de investigación. La más clara es extender el estudio al resto de elementos de la tabla periódica, o al menos de los elementos que sean posibles. Aunque la elaboración del perfil tuvo un buen resultado, del aumento de atributos para cada elemento se espera una mejora de los resultados. Cuanto mayor sea el número de variables de cada perfil, más posibilidades hay de encontrar correlaciones mayores. La desventaja que esto tiene es que los elementos serían cada vez más diferentes, empeorando la llamada clasificación final. Sin embargo optimizar el número de atributos consiguiendo un buen comportamiento de la clasificación de estos, es un problema de gran interés.

Otra alternativa es dar el siguiente paso a la hora de ver cómo afecta la curva DOS a las propiedades finales del material. De esta forma se podría diseñar un material con unas propiedades concretas y a continuación, obtenerlo en el laboratorio y comprobar sus propiedades reales. Lograr buenos resultados en este último proceso, eliminaría una gran cantidad de gastos experimentales, gracias al uso de simulaciones de un coste mucho menor, mejorando considerablemente la producción y eficiencia de los materiales.

9. Bibliografía.

Referencias

[1] Ghanshyam Pilania, Chenchen Wang, Xun Jiang et al *Accelerating materials property predictions using machine learning.* Scientific Reports 2013.

[2] Neil W. Aschroft and N. David Mermin, *Solyd State Physics* Harcourt College Publishers, Printed in united states of America, ISBN 0-03-083993-9

[3] Steven H. Simon *The Oxford Solid State Basics* University Press 2013, ISBN 978-0-19-968076-4.

[4] D.G.Pettifor, *A chemical Scla for ctystal-structure maps.* Department of Mathematics, Imperial College of Science and Technology, London SW7 2BZ, England.