

**UNIVERSIDAD PRIVADA DE CIENCIAS APLICADAS**

Trabajo Final 2019-1

**Integrantes:**

Rodrigo Guadalupe

Nancy Rojas

**Docente:**

Willy Ugarte Rojas

**Curso:**

Administración de la Información

**Carrera:**

Ciencias de la Computación

LIMA – PERU

2019

1. **Ética profesional**
   1. Código de ética profesional del Colegio de Ingenieros del Perú

Es un conjunto de normas y principios fundamentales que guían y regulan el comportamiento de las personas que pertenecen a una empresa u organización. En este caso los profesionales que integren el Colegio de Ingenieros deberán cumplir sus funciones teniendo como base este código de ética.

Aprobado por Unanimidad en la Cuarta Sesión Extraordinaria, realizado en la Ciudad de Lima los días 25, 26, 27 y 28 de junio del 2011 y su modificatoria aprobado por Unanimidad en la Sétima Sesión Extraordinaria, realizado en la Ciudad de Huancayo los días 10 y 11 de agosto del 2012 del Congreso Nacional de Consejos Departamentales del Colegio de Ingenieros del Perú Periodo 2010 – 2012

Con respecto al tratamiento de la información, el Código de ética del Colegio de Ingenieros indica en el capítulo VI – Articulo 45 que el ejercicio de la actividad profesional del Ingeniero implica mantener una conducta decorosa a la vez que honrar los compromisos contraídos, velar por los intereses de los clientes que le encomiendan la ejecución de trabajos inherentes a su profesión.

Por tanto, son deberes del Ingeniero:

a. Servir a su cliente con capacidad, lealtad y empeño profesional. En ningún caso realizará un acto ilícito o contrario a la profesión y atribuyéndolo a instrucciones o imposiciones del cliente.

b. Brindar a sus clientes asesoramiento idóneo y oportuno en las consultas que éstos requieran, recomendándoles aquellas sugerencias técnicas que posibiliten la solución a los problemas planteados.

c. Mantener la debida reserva y discreción respecto a los trabajos, datos e informaciones de índole técnica, financiera y otras de naturaleza confidencial que obtenga de su cliente en el ejercicio de su actividad profesional.

d. Actuar como asesor y defensor de los legítimos intereses del cliente cuando se trate del cumplimiento de contratos entre aquel y terceras personas. Ello no significa que le sea lícito proceder con parcialidad en el ejercicio de estas funciones. Igual conducta aplicará cuando se desempeñe como perito, jurado o árbitro y cuando le correspondiere adjudicar la ejecución de obras o trabajos o la provisión de suministros.

e. Una vez aceptado un encargo no se podrá renunciar a él sino por causa justificada que haya sobrevenido o que sea conocida con posterioridad a la aceptación. Al renunciar debe cuidarse de no perjudicar al cliente.

f. Al descubrir en su trato con el cliente equivocación o impostura de cualquier índole que beneficien a éste injustamente, deberá notificárselo

* 1. Derechos de propiedad intelectual para creadores de software

El Derecho de Autor es un derecho humano reconocido en el artículo 27.2° de la Declaración Universal de Derechos Humanos: “Toda persona tiene derecho a la protección de los intereses morales y materiales que le correspondan por razón de las producciones científicas, literarias o artísticas de que sea autora.”

El régimen sobre el Derecho de Autor está regulado en el Perú por el Decreto Legislativo N° 822 – Ley sobre el Derecho de Autor y contiene una serie de derechos y normas aplicables, en nuestro país, a todos los autores y sus obras.

Los programas informáticos o de software también son considerados “obras” y, de cumplir con los requisitos de originalidad están protegidos por el Derecho de autor. Quedan protegidos los siguientes componentes de un programa:

* El código fuente
* El código objetivo (ejecutable)
* La documentación técnica (memoria descriptiva)
* Los manuales de uso o usuario

1. **Responsabilidad Profesional**
   1. Competencias del perfil profesional de Ingeniería de Software y Ciencias de la Computación
   * Aplicación de conocimientos de ciencias
   * Planificación y conducción de experimentos: Poder usar diversos métodos de planificación, de este modo tener un buen desarrollo del proyecto.
   * Diseño de procesos y sistemas
   * Trabajo en equipos multidisciplinarios: tener en cuenta que en un proyecto no se trabaja siempre con profesionales de la misma rama, de lo contrario, muchos equipos obtienen mejores resultados al tener más diversidad en el grupo.
   * Responsabilidad ética y profesional: seguir las reglas estipuladas en el código de ética de Ingenieros, la IEEE y la ACM.
   * Comunicación efectiva: un optimo trabajo en equipo, en el cual pueda desarrollarse tanto profesional como personalmente.
   * Análisis del impacto de la solución de ingeniería.
   * Resolución de problemas de ingeniería (IS)
   * Aprendizaje continuo y autónomo (IS)
   * Conocimientos Contemporáneos (IS)
   * Uso de herramientas modernas de ingeniería (IS).
   * Necesidad de aprendizaje de por vida (CC)
   * Uso de técnica y herramientas (CC): conseguir o construir nuevas herramientas que puedan facilitar el desarrollo de un nuevo software.
   * Aplicación de fundamentos matemáticos (CC): gran parte de la carrera de ciencias de la computación se basa en fundamentos matemáticos implementados dentro de programas.
   * Diseño y construcción de sistemas complejos (CC).

El análisis, tratamiento y visualización de datos aportan de diferentes modos al perfil profesional de un ingeniero de software.

Para realizar todas estas tareas es necesario planificar cada paso que daremos en el desarrollo de este nuevo programa, saber diseñar los procesos y tener en cuenta el orden de estos para no cometer errores que podrían ser fatales al momento de testear el proyecto.

A fin de garantizar la obtención de un programa en óptimas condiciones, nos podemos valer de diversas herramientas que nos proporcionen, ya sean de Internet o algún otro medio; y se puede tener conocimiento previo de estas herramientas o no. Muchas personas aprenden en la marcha, por esto un profesional perteneciente a la facultad de ingeniería Sistemas y computación debe ser consciente de tener un aprendizaje continuo y riguroso. La tecnología no solo avanza para nosotros, sino que se va creando a partir de mentes brillantes que se atrevieron a descubrir más. Muchos profesionales de diferentes ramas necesitan aprender cosas nuevas durante el tiempo que se mantienen activos, pero nosotros necesitamos crear la tecnología que les va a servir a los demás para su desarrollo.

* 1. IEEE Computer Society

Es una sociedad profesional de IEEE. Su propósito y alcance es "avanzar en la teoría, la práctica y la aplicación de la ciencia y la tecnología de procesamiento de la información y la computadora" y la "posición profesional de sus miembros". La Computer Society (CS) es la más grande de las 39 sociedades técnicas organizadas bajo el IEEE Technical Activities Board (Junta de Actividades Técnicas).

IEEE Computer Society patrocina talleres y conferencias, publica una variedad de publicaciones revisadas por pares, opera comités técnicos y desarrolla estándares de computación IEEE. Participa en actividades educativas en todos los niveles de la profesión, incluido el aprendizaje a distancia, la acreditación de programas de educación superior en informática y la certificación profesional en ingeniería de software.

Es también una organización miembro de la Federación de Organizaciones Profesionales de Arquitectura Empresarial (una asociación mundial de organizaciones profesionales que se han unido para proporcionar un foro para estandarizar, profesionalizar y, por lo demás, promover la disciplina de la Arquitectura Empresarial).

* 1. ACM Code of Ethics and Professional Conduct

**Código de ética y conducta profesional**

El Código está diseñado para inspirar y guiar la conducta ética de todos los profesionales de la informática, incluidos los profesionales, instructores, estudiantes, personas influyentes y personas que utilizan la tecnología informática de manera impactante. Además, el Código sirve como base para la reparación cuando se producen violaciones. El Código incluye principios formulados como declaraciones de responsabilidad, basados ​​en el entendimiento de que el bien público es siempre la consideración principal. Cada principio se complementa con pautas, que brindan explicaciones para ayudar a los profesionales de la computación a comprender y aplicar el principio.

El Código en su conjunto se ocupa de cómo se aplican los principios éticos fundamentales a la conducta de un profesional informático. El Código no es un algoritmo para resolver problemas éticos; más bien sirve como base para la toma de decisiones éticas. Al pensar en un tema en particular, un profesional de la computación puede encontrar que se deben tener en cuenta múltiples principios y que diferentes principios tendrán una relevancia diferente para el problema. Las preguntas relacionadas con este tipo de problemas pueden responderse mejor mediante una consideración cuidadosa de los principios éticos fundamentales, entendiendo que el bien público es la consideración primordial. Toda la profesión informática se beneficia cuando el proceso de toma de decisiones éticas es responsable y transparente para todas las partes interesadas. Las discusiones abiertas sobre cuestiones éticas promueven esta responsabilidad y transparencia.

* **PRINCIPIOS ÉTICOS GENERALES.**

El código de ética ACM Capitulo 1 Articulo 7 - Honrar la confidencialidad.

A los profesionales de la informática a menudo se les confía información confidencial, como secretos comerciales, datos de clientes, estrategias comerciales no públicas, información financiera, datos de investigación, artículos académicos previos a la publicación y solicitudes de patentes. Los profesionales de la computación deben proteger la confidencialidad, excepto en los casos en que sea evidencia de la violación de la ley, de los reglamentos de la organización o del Código. En estos casos, la naturaleza o el contenido de esa información no se debe divulgar excepto a las autoridades apropiadas. Un profesional de la computación debe considerar cuidadosamente si dichas revelaciones son consistentes con el Código.

* **RESPONSABILIDADES PROFESIONALES**

El código de ética ACM Capitulo 2 Articulo 8 - Acceder a los recursos informáticos y de comunicación solo cuando esté autorizado o cuando esté obligado por el bien público.

Los individuos y las organizaciones tienen el derecho de restringir el acceso a sus sistemas y datos siempre que las restricciones sean consistentes con otros principios del Código. En consecuencia, los profesionales de la computación no deben acceder al sistema informático, software o datos de otra persona sin una creencia razonable de que tal acción sería autorizada o una creencia convincente de que es consistente con el bien público. Un sistema de acceso público no es motivo suficiente por sí solo para implicar una autorización. En circunstancias excepcionales, un profesional de la informática puede utilizar el acceso no autorizado para interrumpir o inhibir el funcionamiento de los sistemas maliciosos; Se deben tomar precauciones extraordinarias en estos casos para evitar daños a otros.

* 1. Implementación y desarrollo
     1. Descripción del dataset

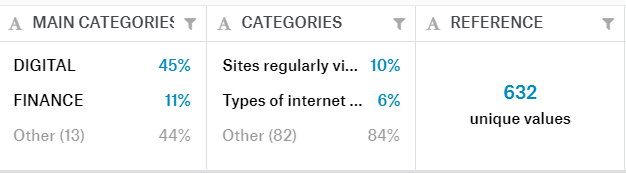
*La base de datos utilizada es “Smart meters in London” y fue obtenida de la página kaggle.com (comunidad en línea de científicos de datos y aprendices de máquinas, propiedad de Google LLC).*

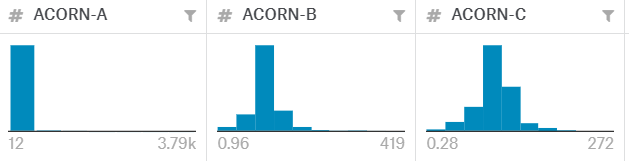
*Todo el gran trabajo de recolección de datos ha sido realizado por las redes eléctricas del Reino Unido para los datos del medidor inteligente*

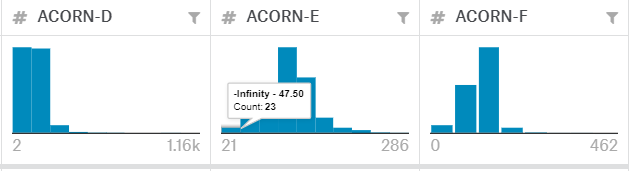
La Unión Europea solicitó a todos los gobiernos miembros que consideren los medidores inteligentes como parte de las medidas para mejorar nuestro suministro de energía y hacer frente al cambio climático. Después de un estudio inicial, el gobierno británico decidió adoptar medidores inteligentes como parte de su plan para actualizar su sistema de energía que está envejeciendo. Para seguir mejor el consumo de energía, el gobierno quiere que los proveedores de energía instalen medidores inteligentes en cada hogar en Inglaterra, Gales y Escocia. Hay más de 26 millones de hogares a los que pueden acceder los proveedores de energía, con el objetivo de que cada hogar tenga un medidor inteligente para 2020.

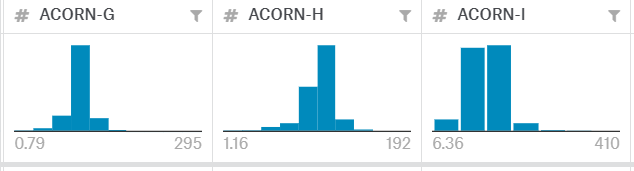
En este conjunto de datos, encontrará una versión refactorizada de los datos del almacén de datos de Londres, que contiene las lecturas de consumo de energía para una muestra de 5,567 hogares de Londres que participaron en el proyecto de Low Carbon London liderado por UK Power Networks entre noviembre de 2011 y febrero. 2014. Los datos de los medidores inteligentes parecen asociados solo al consumo eléctrico.

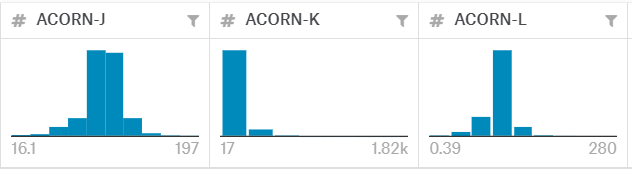
Dimensiones del Dataset

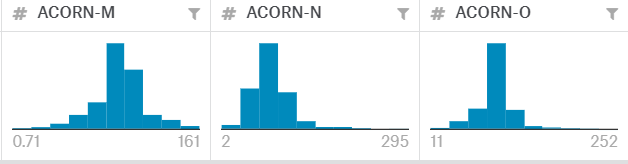












* + 1. Descripción del pre-procesamiento y limpieza del dataset

Normalizacion por Maximin: Las ganancias serán el máximo de los mínimos. Bajo este criterio se analizan los resultados que ofrecen cada una de las situaciones relacionadas con la alternativa (estados de la naturaleza), de estas se escoge la peor condición. Esto se hace para todas las alternativas y entre estas se escoge la alternativa que ofrece los mejores resultados entre los peores.

* + 1. Descripción de los métodos implementados
* Weighted Average
* Maximin/minimax
* Leximin/leximax
* Skylines
  + 1. Optimización de los datos multi-dimensionales
    - Weighted Average: Un promedio ponderado es un tipo de promedio donde cada observación en el conjunto de datos se multiplica por un peso predeterminado antes del cálculo. Al calcular un promedio simple (media aritmética), todas las observaciones se tratan por igual y se les asigna el mismo peso. Un promedio ponderado asigna ponderaciones que determinan la importancia relativa de cada punto de datos. Las ponderaciones son equivalentes a tener tantos elementos similares con el mismo valor involucrado en el promedio.
    - Maximin/minimax: es una regla de decisión utilizada en inteligencia artificial, teoría de decisiones, teoría de juegos, estadísticas y filosofía para minimizar la posible pérdida en el peor de los casos (pérdida máxima). Cuando se trata de ganancias, se le conoce como "maximin", para maximizar la ganancia mínima. Formulado originalmente para la teoría de juegos de suma cero para dos jugadores, que abarca tanto los casos en los que los jugadores toman movimientos alternativos como los que realizan movimientos simultáneos, también se ha extendido a juegos más complejos y a la toma de decisiones en general en presencia de incertidumbre.
    - Leximin/leximax: Funciones objetivas dispuestas en orden de importancia. Existen varias variantes y generalizaciones del ordenamiento lexicográfico. Una variante ampliamente utilizada en combinatoria ordena subconjuntos de un conjunto finito dado asignando un orden total al conjunto finito, y convirtiendo subconjuntos en secuencias crecientes, a las que se aplica el orden lexicográfico.
    - Skylines: Un skyline es una colección de tiras rectangulares. Una tira rectangular se representa como un par (izquierda, ht) donde la izquierda es la coordenada x del lado izquierdo de la tira y ht es la altura de la tira.

"Izquierda": es x coordinada del lado izquierdo (o pared).

"Derecha": es la coordenada x del lado derecho

'Ht': es la altura del edificio.

Dados n edificios rectangulares en una ciudad bidimensional, calcula el horizonte de estos edificios, eliminando líneas ocultas. La tarea principal es ver los edificios desde un lado y eliminar todas las secciones que no son visibles.

Todos los edificios comparten el fondo común y cada edificio está representado por un triplete (izquierda, derecha, derecha)

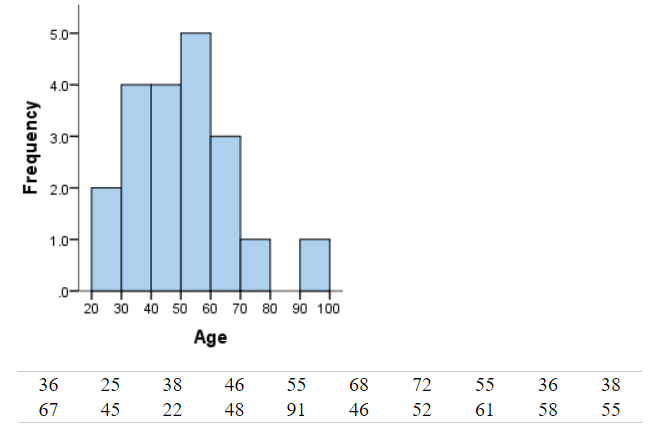
* + 1. Agrupamiento de datos multi-dimensionales
* K-Means: K-Means es un algoritmo [no supervisado](http://www.aprendemachinelearning.com/aplicaciones-del-machine-learning/#no_supervisado) de agrupamiento o [Clustering](http://www.aprendemachinelearning.com/principales-algoritmos-usados-en-machine-learning/#clustering). Se utiliza cuando existen muchos datos sin etiquetar. El objetivo de este algoritmo es el de encontrar «K» grupos (clusters) entre los datos crudos.

El algoritmo trabaja iterativamente para asignar a cada «punto» (las filas de nuestro conjunto de entrada forman una coordenada) uno de los «K» grupos basado en sus características. Son agrupados en base a la similitud de sus features (las columnas). Como resultado de ejecutar el algoritmo tendremos:

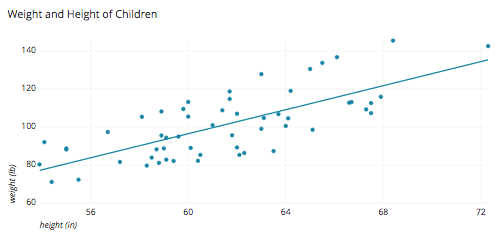
* Los «centroids» de cada grupo que serán unas «coordenadas» de cada uno de los K conjuntos que se utilizarán para poder etiquetar nuevas muestras.
* Etiquetas para el conjunto de datos de entrenamiento. Cada etiqueta perteneciente a uno de los K grupos formados.

Los grupos se van definiendo de manera «orgánica», es decir que se va ajustando su posición en cada iteración del proceso, hasta que converge el algoritmo. Una vez hallados los centroids deberemos analizarlos para ver cuáles son sus características únicas, frente a la de los otros grupos. Estos grupos son las etiquetas que genera el algoritmo.

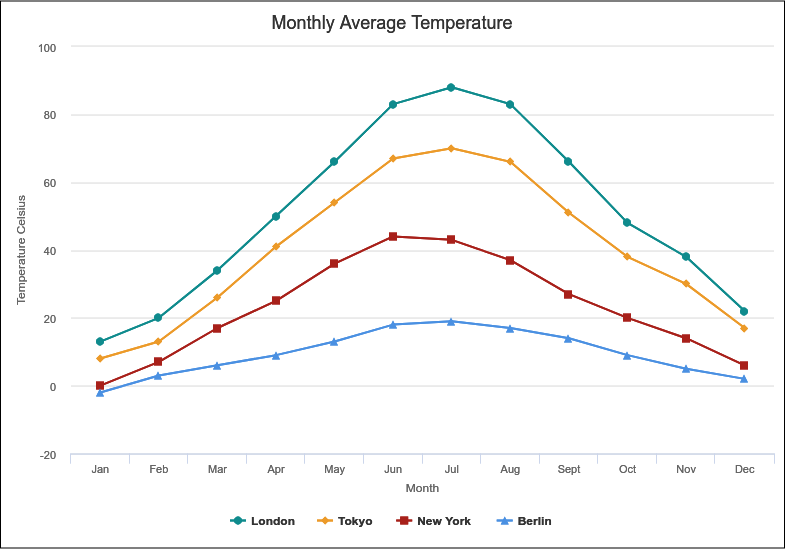
* + 1. Técnica de clasificación
* KNN: K-Nearest-Neighbor es un algoritmo [basado en instancia](http://www.aprendemachinelearning.com/principales-algoritmos-usados-en-machine-learning/#instancia) de tipo [supervisado](http://www.aprendemachinelearning.com/aplicaciones-del-machine-learning/#supervisado) de Machine Learning. Puede usarse para clasificar nuevas muestras (valores discretos) o para predecir (regresión, valores continuos). Al ser un método sencillo, es ideal para introducirse en el mundo del Aprendizaje Automático. Sirve esencialmente para clasificar valores buscando los puntos de datos «más similares» (por cercanía) aprendidos en la etapa de entrenamiento y haciendo conjeturas de nuevos puntos basado en esa clasificación.
  + 1. Reducción de dimensionalidad
* PCA: Principal Component Analysis es una técnica de Extracción de Características donde combinamos las entradas de una manera específica y podemos eliminar algunas de las variables «menos importantes» manteniendo la parte más importante todas las variables. Como valor añadido, luego de aplicar PCA conseguiremos que todas las nuevas variables sean independientes una de otra.
  + 1. Visualización de datos multi-dimensionales
* Histogramas: Un histograma es un gráfico que permite descubrir y mostrar la distribución de frecuencia subyacente (forma) de un conjunto de datos continuos. Esto permite la inspección de los datos para su distribución subyacente (por ejemplo, distribución normal), valores atípicos, asimetría, etc. A continuación, se muestra un ejemplo de un histograma y los datos sin procesar con los que se construyó.



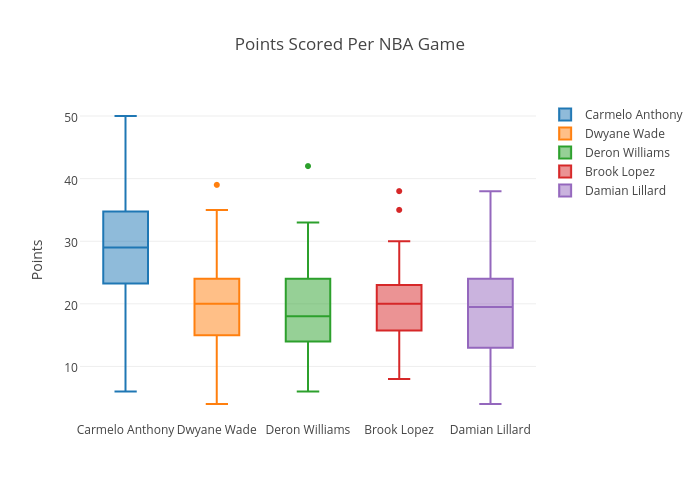
* Scatter Plots: Un diagrama de dispersión es una visualización de datos bidimensionales que utiliza puntos para representar los valores obtenidos para dos variables diferentes: una representada a lo largo del eje x y la otra representada a lo largo del eje y. Por ejemplo, este diagrama de dispersión muestra la altura y el peso de un conjunto de niño.



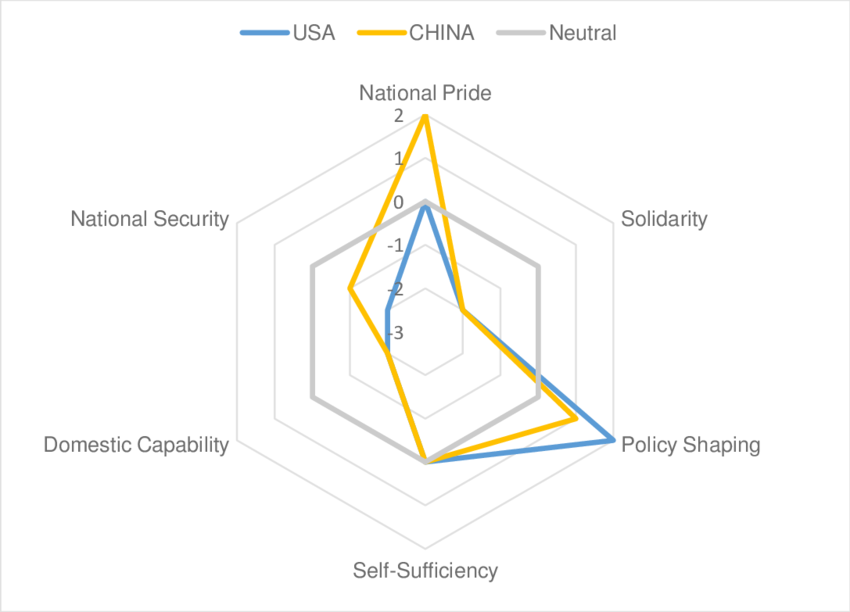
* Line plots: Los gráficos de líneas se utilizan generalmente cuando necesitamos mostrar los datos que se pueden representar como una tabla en una hoja de cálculo, también para mostrar algunos cambios durante un período de tiempo, como por ejemplo con las temperaturas o las tasas y se realizan en un sistema de coordenadas dimensional normal.



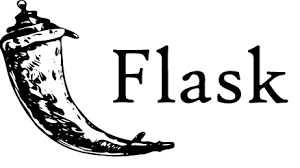
* Box Plot: es una forma estandarizada de mostrar la distribución de datos basada en el resumen de cinco números: mínimo, primer cuartil, mediana, tercer cuartil y máximo.



* Hexagon plot: Este tipo de gráfico es realmente útil si su diagrama de dispersión es demasiado denso para interpretar. Ayuda a agrupar el área espacial de la tabla y la intensidad del color que un hexágono puede interpretar como puntos que se concentran más en esta área.



* + 1. Combinaciones de métodos de Optimización y/o Agrupamiento
* PCA + KNN: se utilizará PCA para reducir las dimensiones, se tomarán las dimensiones más importantes y se procederá a realizar agrupamiento con KNN.
* PCA + K-means: se utilizará PCA para reducir las dimensiones, se tomarán las dimensiones más importantes y se procederá a realizar agrupamiento con K-Means.
  + 1. Descripción del formulario grafico
* Flask: es un framework minimalista escrito en Python que permite crear aplicaciones web rápidamente y con un mínimo número de líneas de código. Esta basado en la especificación WSGI de Werkzeug y el motor de templates Jinja2 y tiene una licencia BSD



1. **Conclusiones**
   1. Resultados relevantes
   2. Buenas prácticas de visualización

Para hacer una buena visualización de datos es de vital importancia disponer de buenos datos. Esto significa que se tendrá que trabajar con los datos antes de representarlos, con el fin de poder mostrar el mensaje que se esconde detrás de ellos. Un claro ejemplo de preparación de datos es el proceso de normalización.

* 1. Impacto Social
  2. Impacto económico