|  |  |
| --- | --- |
| **Instrumento** | *Reporte* |

|  |  |  |
| --- | --- | --- |
| **Alumno**: Raul Ponce Huerta | | **Fecha:** 12/07/2024 |
| **Carrera:** Ingeniería en Desarrollo y Gestión de Software | | **Grupo:** IDGS91 |
| **Asignatura:** Extracción de Conocimiento en Base de Datos | |  | | --- | | **Unidad temática:** IV Análisis no supervisado | | |
| **Profesor:** MGTI. María Eugenia Guerrero Chan | | |

Título: Análisis no supervisado

1. Temas
   1. **Algoritmos de aprendizaje no supervisado**
      1. Que son los Algoritmos de aprendizaje no supervisado y tipos.
      2. Que es el Análisis no supervisado
      3. Usos y objetivo del análisis no supervisado.
      4. Que son los Algoritmos de agrupación y tipos
      5. Ejemplos donde se aplican los Algoritmos de agrupación.
      6. Que son los Algoritmos de reducción de dimensionalidad y tipos.
      7. Ejemplos donde se aplican los Algoritmos de reducción de dimensionalidad.
   2. **Métricas de evaluación de modelos de procesamiento de datos.**
      1. Que son métricas de evaluación de modelos de agrupación y reducción de dimensionalidad.
      2. Que es el proceso de entrenamiento y evaluación de agrupación y reducción de dimensionalidad.
      3. Que es el proceso de optimización de modelos de agrupación y reducción de dimensionalidad.
   3. **Algoritmos de aprendizaje no supervisado**

**1.1.1. ¿Qué son los algoritmos de aprendizaje no supervisado y tipos?**

Los algoritmos de aprendizaje no supervisado son métodos de machine learning que buscan patrones en los datos sin utilizar etiquetas predefinidas. A diferencia del aprendizaje supervisado, donde los algoritmos se entrenan con datos etiquetados, el aprendizaje no supervisado trabaja con datos sin etiquetas y busca estructuras inherentes en el conjunto de datos.

**Tipos de algoritmos de aprendizaje no supervisado:**

* Algoritmos de agrupación (clustering): Agrupan datos similares en conjuntos o clústeres.
* Algoritmos de reducción de dimensionalidad: Reducen el número de variables en un conjunto de datos, manteniendo la mayor cantidad posible de información relevante.
* Modelos de mezcla: Modelan la distribución de los datos como una combinación de distribuciones más simples (ej., modelos de mezcla gaussiana).

**1.1.2. ¿Qué es el análisis no supervisado?**

El análisis no supervisado es el proceso de aplicar algoritmos de aprendizaje no supervisado para descubrir patrones, estructuras o relaciones en los datos sin la guía de etiquetas preexistentes. Este análisis puede revelar información oculta y proporcionar insights sobre la estructura subyacente de los datos.

**1.1.3. Usos y objetivos del análisis no supervisado**

**Usos:**

* Descubrimiento de patrones ocultos en datos no etiquetados.
* Identificación de segmentos o clústeres en datos de clientes.
* Reducción de la dimensionalidad de los datos para visualización y análisis más manejable.
* Detección de anomalías o outliers en conjuntos de datos.

**Objetivos:**

* Encontrar estructuras subyacentes en los datos.
* Simplificar conjuntos de datos complejos.
* Facilitar la visualización y la interpretación de los datos.
* Preparar datos para otros algoritmos de machine learning.

**1.1.4. ¿Qué son los algoritmos de agrupación y tipos?**

Los algoritmos de agrupación (clustering) son técnicas de aprendizaje no supervisado que dividen un conjunto de datos en subgrupos homogéneos llamados clústeres. Los datos dentro de un clúster son más similares entre sí que a los datos de otros clústeres.

**Tipos de algoritmos de agrupación:**

* K-means: Asigna puntos de datos a K clústeres basándose en la distancia mínima a los centroides de los clústeres.
* Jerárquico: Construye una jerarquía de clústeres mediante una serie de fusiones o divisiones.
* DBSCAN (Density-Based Spatial Clustering of Applications with Noise): Agrupa puntos de datos en clústeres basándose en densidades locales.
* Mean Shift: Agrupa puntos de datos desplazando iterativamente los puntos hacia regiones de mayor densidad.

**1.1.5. Ejemplos donde se aplican los algoritmos de agrupación**

* Segmentación de clientes: Agrupar clientes con características similares para campañas de marketing dirigidas.
* Análisis de redes sociales: Identificar comunidades o grupos de usuarios con intereses similares.
* Biología: Clasificar especies o genes con características similares.
* Detección de fraudes: Identificar patrones anómalos en transacciones financieras.

**1.1.6. ¿Qué son los algoritmos de reducción de dimensionalidad y tipos?**

Los algoritmos de reducción de dimensionalidad transforman datos de alta dimensión a un espacio de menor dimensión, conservando la información esencial. Esto facilita la visualización, el almacenamiento y el procesamiento de los datos.

**Tipos de algoritmos de reducción de dimensionalidad:**

* PCA (Análisis de Componentes Principales): Transforma los datos a un nuevo espacio de menor dimensión mediante la combinación lineal de variables originales.
* t-SNE (t-distributed Stochastic Neighbor Embedding): Reduce la dimensionalidad conservando las relaciones locales entre los datos.
* LDA (Análisis Discriminante Lineal): Encuentra una proyección que maximiza la separación entre diferentes clases.
* Autoencoders: Redes neuronales que aprenden una representación comprimida de los datos.

**1.1.7. Ejemplos donde se aplican los algoritmos de reducción de dimensionalidad**

* Visualización de datos: Reducir datos de alta dimensión para crear gráficos en 2D o 3D.
* Preprocesamiento de datos: Reducir el número de características antes de aplicar otros algoritmos de machine learning.
* Análisis de imágenes: Comprimir imágenes manteniendo la calidad visual.
* Compresión de datos: Reducir el tamaño de los datos para almacenamiento eficiente.

**1.2. Métricas de evaluación de modelos de procesamiento de datos**

**1.2.1. ¿Qué son métricas de evaluación de modelos de agrupación y reducción de dimensionalidad?**

Las métricas de evaluación de modelos de agrupación y reducción de dimensionalidad son criterios utilizados para medir la eficacia y la calidad de estos algoritmos.

**Métricas de agrupación:**

* Coeficiente de Silueta: Mide la coherencia interna dentro de los clústeres.
* Índice de Davies-Bouldin: Evalúa la compacidad y la separación entre los clústeres.
* Inercia (para K-means): Mide la suma de las distancias cuadradas de cada punto al centroide de su clúster.

**Métricas de reducción de dimensionalidad:**

* Varianza explicada (para PCA): Proporción de la varianza total que se conserva en las componentes principales.
* Preservación de la distancia (para t-SNE): Evalúa cuánto se conservan las distancias relativas entre los puntos en el espacio reducido.

**1.2.2. ¿Qué es el proceso de entrenamiento y evaluación de agrupación y reducción de dimensionalidad?**

El proceso de entrenamiento y evaluación de agrupación y reducción de dimensionalidad involucra varias etapas:

* Preprocesamiento de datos: Limpieza y normalización de los datos.
* Aplicación del algoritmo: Ejecutar el algoritmo de agrupación o reducción de dimensionalidad en los datos.
* Evaluación: Usar métricas específicas para evaluar la calidad del modelo.
* Interpretación: Analizar los resultados para obtener insights significativos.

**1.2.3. ¿Qué es el proceso de optimización de modelos de agrupación y reducción de dimensionalidad?**

El proceso de optimización implica ajustar los parámetros del modelo para mejorar su rendimiento. Incluye:

* Selección de parámetros: Elegir los valores adecuados para los hiperparámetros del modelo (ej., número de clústeres en K-means).
* Validación cruzada: Dividir los datos en conjuntos de entrenamiento y validación para evaluar la robustez del modelo.
* Evaluación continua: Usar métricas de evaluación para medir el desempeño a medida que se ajustan los parámetros.
* Ajuste fino: Refinar los parámetros basándose en los resultados de evaluación para obtener el mejor rendimiento posible.

**Bibliografía**

Scikit-learn documentation. (n.d.). Retrieved from <https://scikit-learn.org/stable/documentation.html>

Towards Data Science. (n.d.). Retrieved from <https://towardsdatascience.com/>

KDnuggets. (n.d.). Retrieved from <https://www.kdnuggets.com/>

Coursera. (n.d.). Machine Learning by Andrew Ng. Retrieved from <https://www.coursera.org/learn/machine-learning>

edX. (n.d.). Data Science: Machine Learning by Harvard University. Retrieved from <https://www.edx.org/course/data-science-machine-learning>