

TRANSFORMACIÓN Y NORMALIZACIÓN

Autor: Carlos Abad, martin.abad2016@uteq.edu.ec

5 de mayo de 2021

Universidad Técnica Estatal de Quevedo

Facultad de Ciencias de la Ingeniería

Quevedo, Ecuador

Resumen: En el presente trabajo abordaré sobre “Transformación y Normalización”, el cual comprende la Discretización, Métodos de Normalización y un caso práctico que comprende 3 análisis de correspondencia (hombres, mujeres y datos conjuntos).

Se utilizó el conjunto de datos Basketball, mismo que fue facilitado por el Ing. Iván Jaramillo, se aplicó los métodos de normalización más utilizados, entre los que podemos destacar, Normalización Z-SCORE, Min-Max, por escalamiento decimal, sigmoidal y Softmax.

I. INTRODUCCIÓN

La necesidad de procesar y extraer conocimiento valioso de tal inmensidad de datos se ha convertido en un desafío considerable para científicos de datos y expertos en la materia [1]. El valor del conocimiento extraído es uno de los aspectos esenciales de Big Data, como la discretización y normalización que son encargados de transformar atributos continuos usando intervalos discretos, mientras que la normalización realiza un ajuste en la distribución, y el análisis de correspondencia principal (PCA) con análisis de correspondencia (CA) reducen la dimensión del problema [2].

Para la elaboración de este trabajo en equipo, se realizó una revisión de los temas de la semana 6 en la cual se ha hecho una investigación profunda, misma se ha compartido en dialogo con el equipo de trabajo en varias sesiones de Google Meet y a su vez se ha realizado los respectivos trabajos de cada semana.

En este trabajo se investigó sobre las diferencias y similitudes de los métodos que tienen los softwares (Weka, R, Knime), Los diferentes métodos de normalización y el análisis de correspondencia para conjuntos de datos.

II. ÍNDICE DE ACTIVIDADES

1. Trabajo 1: Transformación y Normalización.

Fecha: miércoles 16 de junio de 2021

Trabajo: Grupal

La normalización de un conjunto de datos consiste en ajustar los valores medidos en diferentes escalas

respecto a una escala común, es decir, para reducir su redundancia. Para ello se aplican diversas técnicas de normalización entre la más conocida Z-SCORE.

III. METODOLOGÍA / PROCEDIMIENTO

1. Procedimiento del trabajo 1: Transformación y Normalización.

Fecha: miércoles 16 de junio de 2021

Trabajo: Grupal

Este trabajo se llevó a cabo con el grupo mediante una modalidad virtual a través de la aplicación de WhatsApp, Google Meet y la nueva herramienta de Microsoft 365 (OneDrive) proporcionado por la UTEQ, nos pusimos de acuerdo para realizar los stops y el caso práctico y luego se procedió a realizar una investigación individual sobre normalización para después discutir sobre lo investigado con ejemplos y se procedió a realizar la práctica a través de Google Meet.

IV. RESULTADOS

1. Resultado del trabajo 1: Transformación y Normalización.

Fecha: miércoles 16 de junio de 2021

Trabajo: Grupal

Como resultado del Stop1, se realizó un cuadro de diferencias y similitudes de las diversas técnicas o métodos los programas Weka, R, Knime, luego para el Stop2 se utilizó el conjunto de datos Basketball para realizar los diferentes métodos de normalización, y para el caso práctico se realizó los respectivos análisis de correspondencia para hombres, mujeres y datos conjuntos, utilizando el conjunto de datos “HairEyeColor” el cual viene por defecto en “R”.

V. CONCLUSIÓN

1. Conclusión del trabajo 1: Transformación y Normalización.

Fecha: miércoles 16 de junio de 2021

Trabajo: Grupal

Las técnicas de discretización, normalización y análisis de correspondencia permiten tener una visión más clara de los conjuntos de datos tratados.

REFERENCES

- [1] S. Agarwal, *Data mining: Data mining concepts and techniques*. 2014.
- [2] R. Sowmya y K. R. Suneetha, «Data Mining with Big Data», *Proc. 2017 11th Int. Conf. Intell. Syst. Control. ISCO 2017*, vol. 26, n.º 1, pp. 246-250, 2017, doi: 10.1109/ISCO.2017.7855990.

Enlaces de Google Drive:

Trabajo 1 – Formal IEEE:

<https://drive.google.com/file/d/1-OUrhZq30nYGzc0WbZg-Y1Ulkbf8wW5/view?usp=sharing>

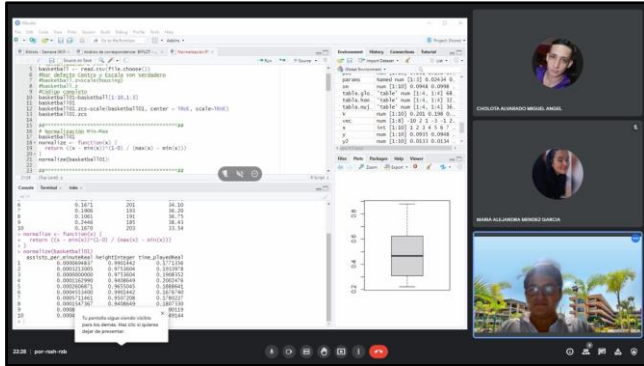
Trabajo 1 – Grupal:

<https://drive.google.com/file/d/1cZs1h36y0rTxBIUXrUh-NUn4EGrtB1cu/view?usp=sharing>

Carpeta del estudiante Semana 6:

https://drive.google.com/drive/folders/1_2PFD4gOa0MGvF9cT5WG92BBcEfrF1b?usp=sharing

ANEXOS



Anexo I. Diálogo y resultado de Normalización de escalamiento decimal.

Tabla 1: Tabla de Actividades designadas

Tarea asignada	Delegado/s
Revisión bibliográfica	Carlos Abad
Redacción de la Introducción	Carlos Abad
Redacción de la Revisión Literaria	Carlos Abad
Cuadro de diferencias y similitudes de los métodos los programas Weka, R, Knime	Alejandra Méndez
Elaboración del resumen y conclusión	Todos
Edición del formato y corrección	Carlos Abad
Diferentes métodos de Normalización	Carlos Abad Miguel Cholota