**PROCESO DE GESTIÓN DE FORMACIÓN PROFESIONAL INTEGRAL**

**TALLER DE CIENCIA DE DATOS – ENTENDIMIENTO DE LOS DATOS**

**IDENTIFICACIÓN DE LA GUIA DE APRENDIZAJE**

* **Denominación del Programa de Formación:** Tecnólogo en Análisis y Desarrollo de Software
* **Código del Programa de Formación:** 228118
* **Nombre del Proyecto:** 2417781 - Desarrollo de software como recurso para la gestión de información de las organizaciones en el contexto de la innovación, transformación digital y en el marco de políticas y requisitos de la empresa.
* **Fase del Proyecto:** Evaluación
* **Actividad de Proyecto:** Desarrollar las tareas de configuración y puesta en marcha del software
* **Competencia:** Construcción del software
* **Resultados de Aprendizaje Alcanzar:** Codificar la solución que cumpla con el diseño establecido.
* **Duración de la Guía:** 120 horas

**2. DESARROLLO DE LA ACTIVIDAD**

Aplicar las fases de entendimiento del negocio y de los datos para la información que se encuentra en el siguiente repositorio: <https://github.com/majash29/CienciaDatosSENA/tree/a7e463f8b0e066d369159f6d253354c36aa5bf9c/TallerEntendimientoDatos>

**Proceso a seguir:**

1. Integración de los datos en un solo dataset.
2. Eliminar variables irrelevantes y redundantes
3. Descripción estadística de los datos
4. Limpieza de datos
5. Transformación de tipo de datos según el método

**3. REFERENTES BILBIOGRÁFICOS**

1. *McKinsey Global Institute. (2018). Notes from the AI frontier: Insights from hundreds of use cases. Recuperado el 22 de marzo de 2023, de* [*https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning#*](https://www.mckinsey.com/featured-insights/artificial-intelligence/notes-from-the-ai-frontier-applications-and-value-of-deep-learning)
2. *Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.*
3. *Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT press.*
4. *James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. Springer.*
5. *Bishop, C. M. (2006). Pattern recognition and machine learning. Springer.*
6. *Shmueli, G. (2010). To explain or to predict?. Statistical science, 25(3), 289-310.*
7. *Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: trends, perspectives, and prospects. science, 349(6245), 255-260.*
8. *Kelleher, J. D., Tierney, B., & Tierney, B. (2018). Data science an introduction. CRC Press.*
9. *LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.*
10. *VanderPlas, J. (2016). Python data science handbook: Essential tools for working with data. O'Reilly Media, Inc.*
11. *Hastie, T., Tibshirani, R., & Wainwright, M. (2015). Statistical learning with sparsity: the lasso and generalizations. CRC Press.*

**7. CONTROL DEL DOCUMENTO**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Nombre** | **Cargo** | **Dependencia** | **Fecha** |
| **Autor (es)** | Alvaro Pérez Niño | Instructor | ADSO | 23 de Abril de 2023 |