

**FORMATO DE GUIÓN INSTRUCCIONAL ACTUALIZACIÓN DE CONTENIDO**

**MAESTRÍA EN CIENCIA DE DATOS E INFORMACIÓN**

|  |  |
| --- | --- |
| **NOMBRE DEL CURSO** | Seminario de Proyectos II |
| **CICLO** | 3 |

**El presente formato tiene el objetivo de registrar las actualizaciones necesarias del contenido a partir de la revisión de la asignatura en aula virtual.**

|  |  |
| --- | --- |
| **VISUALIZACIÓN EN PLATAFORMA** |  |
| **PRESENTACIÓN**  **Instrucciones:**  **Responda las siguientes preguntas para integrar la infografía de presentación del curso, cada respuesta debe tener una extensión de máximo 6 renglones y las respuestas deben ser breves y claras.**   |  |  | | --- | --- | | **Pregunta** | **Respuesta** | | **¿De qué trata la asignatura?** | El desarrollo de esta asignatura te permitirá dar continuidad al proyecto de investigación recepcional. Te enfocarás en el preprocesamiento y análisis de su base de datos, la elección de algoritmos y ajuste de parámetros. | | **¿Por qué es importante en la formación del estudiante?** | Fortalecerás el área de análisis exploratorio de datos que proporciona las herramientas suficientes para determinar las características de los algoritmos más útiles abordados en el problema de investigación. | | **¿Cómo aplica los conocimientos el estudiante en el campo laboral?** | Al finalizar tus estudios,estarás apto para aplicar las herramientas de preprocesamiento, análisis exploratorio, elección y ajuste de algoritmos a problemáticas específicas con bases de datos previamente construidas.  Es momento de que profundices tus conocimientos en el **Seminario de proyectos II.**  ¡Te deseamos mucho éxito! | |  |
| **OBJETIVO GENERAL**  Definir los alcances, límites y justificación del trabajo de investigación así como realizar el preprocesamiento y análisis de la base de datos para elegir y ajustar los modelos de Ciencia de Datos que darán respuesta a las hipótesis de investigación. |  |
| **RUTA DE APRENDIZAJE**   1. Límites, alcances y justificación    1. Identificación de los límites de una investigación    2. Identificación de los alcances de una investigación    3. Redacción de la justificación de un proyecto 2. Preprocesamiento de la base de datos    1. Características de una base de datos que la hacen candidata a un proceso de limpieza    2. Técnicas de preprocesamiento de bases de datos    3. Identificación y aplicación de las técnicas de limpieza de bases de datos 3. Análisis exploratorio de los datos    1. Tipos de variables    2. Gráficos univariantes    3. Gráficos multivariantes    4. Presentación de resúmenes estadístics    5. El Dashboard 4. Elección y ajuste de los modelos    1. Tipos de modelos de la Ciencia de Datos    2. Aplicaciones de los modelos de la Ciencia de Datos    3. Elección de un modelo de la Ciencia de Datos para abordar una problemática de la vida real    4. Aplicación de los modelos seleccionados |  |
| **AGENDA**  Ponemos a tu disposición la agenda de las actividades a realizar durante este **semestre.**   |  |  |  |  |  | | --- | --- | --- | --- | --- | | **UNIDAD** | **ACTIVIDAD**  **(incluir las videoconferencias)** | **TIPO DE ACTIVIDAD** | **Fecha de entrega (Fecha LÍMITE)** | **Puntaje** | | 1 | **Videoconferencia** |  | 17 de enero del 2023 de 18:00 a 19:00 hrs. |  | | 1A. Ejercicio: Límites y Alcances | Individual | 24 de enero del 2023 | 5 | | 1B. Ejercicio: Justificación I | Individual | 7 de febrero del 2023 | 5 | | 1C. Ejercicio: Justificación II | Individual | 14 de febrero del 2023 | 5 | | 2 | **Videoconferencia** |  | 14 de febrero del 2023 de 18:00 a 19:00 hrs |  | | 2A. Resumen: Motivación del preprocesamiento | Individual | 28 de febrero del 2023 | 5 | | 2B. Reporte: Aplicación de las herramientas de preprocesamiento | Individual | 14 de marzo del 2023 | 15 | | 3 | **Videoconferencia** |  | 14 de marzo del 2023 de 18:00 a 19:00 hrs |  | | 3A. Resumen: Análisis exploratorio de datos | Individual | 28 de marzo del 2023 | 5 | | 3B. Infografía: Dashboard | Individual | 18 de abril del 2023 | 10 | | 4 | **Videoconferencia** |  | 18 de abril del 2023 de 18:00 a 19:00 hrs |  | | 4A. Resumen: Elección de los modelos de Ciencia de Datos | Individual | 2 de mayo del 2023 | 10 | | 4B.Presentación: Ajuste de los modelos de Ciencia de Datos | Individual | 16 de mayo del 2023 | 40 | | **TOTAL**  **Recuerde que la suma de las actividades no debe rebasar este total** | | | | **100** | |  |

|  |  |
| --- | --- |
| **UNIDAD 1: Límites, Alcances y Justificación** |  |
| **Objetivo**  Continuar con el desarrollo del cuerpo metodológico del proyecto de investigación planteado en la materia de Seminario de proyectos I definiendo los límites y alcances de la propuesta así como la justificación de la realización del mismo. |  |
| **Recursos básicos**  Consulta el siguiente material y tómate el tiempo para entenderlo. Te sugerimos elaborar notas, síntesis o mapas conceptuales de los puntos más importantes.   1. **Metodología de la investigación**   Hernández Sampieri, R., Fernández Collado, C. y Baptista Lucio, P. (2014). Metodología de la Investigación. (6ª ed.). pp. 34-57. México: McGraw Hill.   1. **Metodología de la investigación**   Baena Paz, G. (2017). Metodología de la Investigación. (3ª ed.). pp. 56 y 57. México: Patria   1. **Metodología de la investigación**   Lerma González, H. D. (2009). Metodología de la investigación (4ª ed.). pp. 29-34. Bogotá, Colombia: Ecoe Ediciones.  **Insertar Banner** | **Vincular a:**  *1.Sampieri\_Metodologia.pdf*  *2.Baena\_Metodologia.pdf*  *3.Lerma\_Metodologia.pdf*  ***Banners Posgrados notas Recursos básicos*** |
| **1A. Ejercicio. Límites y Alcances**  **Tipo de actividad: Individual**  **Instrucciones:**   1. **Lee** la página 39 del libro 1 y la página 52 del libro 2 de los Recursos Básicos, referentes a las limitaciones que debe de considerar el investigador. 2. **Lee** la página 90 del libro 1 de Recursos Básico, referente a los alcances de una investigación. 3. A partir de lo aprendido y con base en el proyecto de investigación, **identifica** 3 límites y 3 alcances de tu trabajo. 4. **Escribe** tus hallazgos en un archivo. 5. **Guarda** tu archivo en pdf con el número de la unidad, tipo de actividad (A, B, o C), nombre y apellido de la siguiente manera: U#\_A\_Nombre\_Apellido.pdf 6. **Entrega** tu actividad en el espacio correspondiente.   **Insertar Banner**  **Insertar espacio de entrega 📥** | ***Banners Posgrados notas Actividades*** |
| **1B. Ejercicio. Justificación I**  **Tipo de actividad: Individual**  **Instrucciones:**   1. **Lee** la página 40 del libro 1, la página 59 del libro 2 y la página 38 del libro 3 de los Recursos Básicos, referentes a la Justificación. 2. A partir de lo aprendido y con base en tu trabajo de investigación, **redacta** una justificación de la realización de tu trabajo de investigación. 3. **Guarda** tu archivo en pdf con el número de la unidad, tipo de actividad (A, B, o C), nombre y apellido de la siguiente manera: U#\_A\_Nombre\_Apellido.doc 4. **Entrega** tu actividad en el espacio correspondiente.   **Insertar Banner**  **Insertar espacio de entrega 📥** | ***Banners Posgrados notas Actividades*** |
| **1C. Ejercicio. Justificación II**  **Tipo de actividad: Individual**  **Instrucciones:**   1. **Considera** y **analiza** las observaciones que se te indicaron en la entrega de tu actividad 1B. 2. A partir de dichas observaciones **realiza** los cambios necesarios en tu justificación para perfeccionar la comprensión de esta sección. 3. **Guarda** tu archivo en pdf con el número de la unidad, tipo de actividad (A, B, o C), nombre y apellido de la siguiente manera: U#\_A\_Nombre\_Apellido.doc 4. **Entrega** tu actividad en el espacio correspondiente.   **Insertar Banner**  **Insertar espacio de entrega 📥** | ***Banners Posgrados notas Actividades*** |

|  |  |
| --- | --- |
| RECURSOS COMPLEMENTARIOS |  |
| Si deseas profundizar en los temas revisados, te sugerimos consultar:   1. Bernal, C. A. (2010). Metodología de la Investigación. (3a ed.). Colombia: Pearson. Enlace: <https://abacoenred.com/wp-content/uploads/2019/02/El-proyecto-de-investigaci%C3%B3n-F.G.-Arias-2012-pdf.pdf> 2. Cortés Cortés, M. E. e Iglesias León, M. (2004). Generalidades sobre Metodología de la Investigación. Enlace: <https://www.unacar.mx/contenido/gaceta/ediciones/metodologia_investigacion.pdf> 3. Artiles Visbal, L., Otero Iglesias, J. y Barrios Osuna, I. (2020). Metodología de la Investigación. Enlace: <https://instituciones.sld.cu/ihi/metodologia-de-la-investigacion/> | **Vincular a:**  ***1.***<https://abacoenred.com/wp-content/uploads/2019/02/El-proyecto-de-investigaci%C3%B3n-F.G.-Arias-2012-pdf.pdf>  *2.*<https://www.unacar.mx/contenido/gaceta/ediciones/metodologia_investigacion.pdf>  *3.*<https://instituciones.sld.cu/ihi/metodologia-de-la-investigacion/> |

|  |  |
| --- | --- |
| **UNIDAD 2: Preprocesamiento de la base de datos** |  |
| **Objetivo**  Identificar los atributos de la base de datos de su proyecto de investigación que necesitan un proceso previo a su utilización en los algoritmos de Ciencia de datos así como seleccionar y aplicar las técnicas adecuadas para corregir y estandarizar dicha base de datos. |  |
| **Recursos básicos**  Consulta el siguiente material y tómate el tiempo para entenderlo. Te sugerimos elaborar notas, síntesis o mapas conceptuales de los puntos más importantes.   1. **Big Data: Preprocesamiento y calidad de datos**   García, S., Ramírez-Gallego, S., Luengo, J., & Herrera, F. C. (2016). Big Data: Preprocesamiento y calidad de datos. Novática, 237, 17-23.   1. **Best practices in data cleaning**   Osborne, J. W. (2013). *Best practices in data cleaning: A complete guide to everything you need to do before and after collecting your data*. Sage.   1. **Bad data handbook**   McCallum, Q. E. (2012). *Bad data handbook: cleaning up the data so you can get back to work*. " O'Reilly Media, Inc.".  **Insertar Banner** | **Vincular a:**  *1.BigData\_Procesamiento.pdf*  *2.BestPractices\_Data.pdf*  *3.BadData.pdf*  ***Banners Posgrados notas Recursos básicos*** |
| **2A. Resumen. Motivación del preprocesamiento**  **Tipo de actividad: Individual**  **Instrucciones:**   1. **Revisa** la lectura 1 de los Recursos básicos. 2. **Lee** el capítulo 1 del libro 2 y del libro 3 de los Recursos básicos. 3. A partir de lo aprendido, **reflexiona** y **analiza** las necesidades de preprocesamiento que posee la base de datos de tu investigación. 4. **Realiza** un resumen en donde menciones cada una de las características de tu base de datos que la hacen candidata para aplicar alguna de las técnicas de preprocesamiento estudiadas. 5. **Guarda** tu archivo en pdf con el número de la unidad, tipo de actividad (A, B, o C), nombre y apellido de la siguiente manera: U#\_A\_Nombre\_Apellido.doc 6. **Entrega** tu actividad en el espacio correspondiente.   **Insertar Banner**  **Insertar espacio de entrega 📥** | ***Banners Posgrados notas Actividades*** |
| **2B. Reporte. Aplicación de las herramientas de preprocesamiento**  **Tipo de actividad: Individual**  **Instrucciones:**   1. A partir de las necesidades de preprocesamiento de la base de datos detectadas en la actividad 2A, **implementa** los algoritmos necesarios para subsanar las irregularidades en sus datos. 2. **Crea** un reporte en donde indiques para cada necesidad abordada:    * Característica de la variable/base por la cual fue necesaria la intervención    * Algoritmo o estrategia utilizada    * Estadísticos descriptivos y/o gráficos que muestren el antes y después del procesamiento 3. **Guarda** tu archivo en pdf con el número de la unidad, tipo de actividad (A, B, o C), nombre y apellido de la siguiente manera: U#\_A\_Nombre\_Apellido.doc 4. **Entrega** tu actividad en el espacio correspondiente.   **Insertar Banner**  **Insertar espacio de entrega 📥** |  |

|  |  |
| --- | --- |
| RECURSOS COMPLEMENTARIOS |  |
| Si deseas profundizar en los temas revisados, te sugerimos consultar:   1. Ganti, V., & Sarma, A. D. (2013). Data cleaning: A practical perspective. *Synthesis Lectures on Data Management*, *5*(3), 1-85   Enlace <https://www.morganclaypool.com/doi/abs/10.2200/S00523ED1V01Y201307DTM036>   1. Gudivada, V., Apon, A., & Ding, J. (2017). Data quality considerations for big data and machine learning: Going beyond data cleaning and transformations. *International Journal on Advances in Software*, *10*(1), 1-20.   Enlace <https://www.researchgate.net/profile/Junhua-Ding/publication/318432363_Data_Quality_Considerations_for_Big_Data_and_Machine_Learning_Going_Beyond_Data_Cleaning_and_Transformations/links/59ded28b0f7e9bcfab244bdf/Data-Quality-Considerations-for-Big-Data-and-Machine-Learning-Going-Beyond-Data-Cleaning-and-Transformations.pdf>   1. Hernández, C., & Rodríguez, J. E. R. (2008). Preprocesamiento de datos estructurados. *Revista vínculos*, *4*(2), 27-48   Enlace <https://revistas.udistrital.edu.co/index.php/vinculos/article/download/4123/5790/> | *Formato APA* |

|  |  |
| --- | --- |
| **UNIDAD 3: Análisis exploratorio de datos** |  |
| **Objetivo**  Conocer las diferentes herramientas de estadística descriptiva que sirven para representar estructuras y patrones que existen dentro de los datos con el fin de aplicarlas a la base de datos de su proyecto de investigación y presentar un resumen en formato de dashboard. |  |
| **Recursos básicos**  Consulta el siguiente material y tómate el tiempo para entenderlo. Te sugerimos elaborar notas, síntesis o mapas conceptuales de los puntos más importantes.   1. **Guía práctica de introducción al Análisis Exploratorio de Datos**   Gobierno de España (2021) Guía práctica de introducción al Análisis Exploratorio de Datos.   1. **Análisis descriptivo y exploratorio de datos**   Rojo, J. M. (2006). Análisis descriptivo y exploratorio de datos. *Laboratorio de Estadística del Instituto de Economía y Geografía Consejo Superior de Investigaciones Científicas, Madrid*.   1. **Effective dashboard design**   Janes, A., Sillitti, A., & Succi, G. (2013). Effective dashboard design. *Cutter IT Journal*, *26*(1), 17-24.  **Insertar Banner** | **Vincular a:**  *1.Guia\_IntroduccionAED.pdf*  *2.AnalisisDED.pdf*  *3.Effective\_Dashboard.pdf*  ***Banners Posgrados notas Recursos básicos*** |
| **3A. Resumen. Análisis exploratorio de datos**  **Tipo de actividad: Individual**  **Instrucciones:**   1. **Revisa** la lectura 1 y 2 de los Recursos básicos. 2. A partir de lo aprendido, **reflexiona** y **analiza** las características, patrones y/o relaciones presentes en tu base de datos que pueden ser factibles a ser estudiadas bajo un análisis exploratorio de datos. 3. De las opciones identificadas en el punto 2, **elige** 3 análisis que pueden ser de más ayuda a la hora de implementar los algoritmos de Ciencia de datos a tu base. 4. En un archivo **escribe** para cada uno de los atributos elegidos:    * La(s) características a estudiar    * El tipo de herramienta estadística a utilizar (tabla, estadístico, gráfico) 5. **Guarda** tu archivo en pdf con el número de la unidad, tipo de actividad (A, B, o C), nombre y apellido de la siguiente manera: U#\_A\_Nombre\_Apellido.doc 6. **Entrega** tu actividad en el espacio correspondiente.   **Insertar Banner**  **Insertar espacio de entrega 📥** | ***Banners Posgrados notas Actividades*** |
| **3B. Ejercicio. Dashboard**  **Tipo de actividad: Individual**  **Instrucciones:**   1. **Investiga** qué es y para qué se utiliza un Dashboard. 2. **Busca** ejemplos de Dashboards aplicados a diferentes conjuntos de datos. 3. **Revisa** la lectura 3 de los Recursos básicos. 4. A partir de lo aprendido y considerando la propuesta realizada en la actividad 3A, **realiza** un Dashboard de tu base de datos que incluya por lo menos el análisis de 5 características/facetas diferentes de la información de tu proyecto. 5. **Guarda** tu archivo en pdf con el número de la unidad, tipo de actividad (A, B, o C), nombre y apellido de la siguiente manera: U#\_A\_Nombre\_Apellido.doc 6. **Entrega** tu actividad en el espacio correspondiente.   **Insertar Banner**  **Insertar espacio de entrega 📥** |  |

|  |  |
| --- | --- |
| RECURSOS COMPLEMENTARIOS |  |
| Si deseas profundizar en los temas revisados, te sugerimos consultar:   1. **Fundamentals of data visualization**   Wilke, C. O. (2019). *Fundamentals of data visualization: a primer on making informative and compelling figures*. O'Reilly Media.   1. **Handbook of data visualization**   Chen, C. H., Härdle, W. K., & Unwin, A. (Eds.). (2007). *Handbook of data visualization*. Springer Science & Business Media.   1. **Scientific visualization**   Bonneau, G. P., Ertl, T., & Nielson, G. M. (2006). *Scientific visualization: The visual extraction of knowledge from data* (Vol. 1). Berlin: Springer. | **Vincular a:**  *1.Fundamentals\_DataV.pdf*  *2.Handbook\_DataV.pdf*  *3.Scientific\_Visualization.pdf* |

|  |  |
| --- | --- |
| **UNIDAD 4: Elección y ajuste de los modelos** |  |
| **Objetivo**  Conocer los diferentes algoritmos que la Ciencia de datos utiliza en el campo de aplicación con el fin de elegir, aplicar y ajustar los modelos específicos que sirvan para abordar la problemática planteada en el proyecto de investigación.l |  |
| **Recursos básicos**  Consulta el siguiente material y tómate el tiempo para entenderlo. Te sugerimos elaborar notas, síntesis o mapas conceptuales de los puntos más importantes.   1. **Data Science Algorithms in a Week**   Natingga, D. (2017). *Data Science Algorithms in a Week: Top 7 algorithms for computing, data analysis, and machine learning*. Packt Publishing.   1. **A quick review of machine learning algorithms**   Ray, S. (2019, February). A quick review of machine learning algorithms. In *2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon)* (pp. 35-39). IEEE.   1. **Machine learning algorithms-a review.**   Mahesh, B. (2020). Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, *9*, 381-386.   1. **Machine Learning Algorithms and applications**   Mohammed, M., Khan, M. B., & Bashier, E. B. M. (2016). *Machine learning: algorithms and applications*. Crc Press. | *Formato APA* |
| **4A. Resumen: Elección de los modelos de Ciencia de Datos**  **Tipo de actividad: Individual**  **Instrucciones:**   1. Revise las lecturas 2 y 3 de los Recursos básicos. 2. A partir de lo aprendido, identifique el tipo de algoritmo a utilizar en tu proyecto de investigación. 3. Genere un resumen en donde indique claramente los algoritmos elegidos así como una breve justificación de dicha elección. 4. Exporte su trabajo final a formato pdf   Criterios de evaluación: Indique los criterios de evaluación en la siguiente tabla:   |  |  | | --- | --- | | **Criterio** | **Puntaje** | | Identifica los algoritmos a utilizar en su proyecto | 3 | | Describe de manera puntual la razón por la que utilizará cada uno de los algortimos mencionados | 5 | | La redacción es correcta, clara, concisa y sin errores ortográficos. | 2 | | **Total** | **10** | |  |
| **4B. Presentación: Ajuste de los modelos de Ciencia Datos**  **Tipo de actividad: Individual**  **Instrucciones:**   1. Con apoyo de su asesor y con base en la propuesta de la actividad 4A, realice el ajuste de los parámetros pertinente para su base de datos y algoritmos. 2. Elabore una presentación del avance de proyecto que incluya:    1. Descripción del problema    2. Objetivos    3. Metodología de la construcción de la base de datos    4. Proceso de limpieza de la base de datos    5. Análisis exploratorio de la base de datos (Dashboard)    6. Selección de algoritmos    7. Ajuste de los parámetros de los algoritmos 3. La presentación deberá subirse en cualquier formato (.ppt, .pptx, .pdf u otro) y estar planeada para exponerla en una sesión de 15 minutos ya sea utilizando diapositivas para un presentación en vivo o preparando un video. Las sesiones se llevarán a cabo (tentativamente) los días 23, 24, 25 y 26 de mayo de 6 a 8 pm. 4. Manténgase al pendiente de la calendarización de las presentaciones ya que la calificación de esta actividad se basa en dicha exposición.  |  |  | | --- | --- | | **Criterio** | **Puntaje** | | La definición del tema y los objetivos son claros | 5 | | Explica la fuente de información para su base de datos así como todo el proceso de limpieza de la misma | 5 | | El análisis exploratorio de datos proporciona información relevante sobre los patrones de comportamiento de los datos | 10 | | La selección y ajuste de los algoritmos es adecuada | 10 | | Da respuesta oportuna y concisa a las preguntas de los examinadores | 5 | | Se ajusta al tiempo indicado para la presentación | 5 | | **Total** | **40** | |  |

|  |  |
| --- | --- |
| RECURSOS COMPLEMENTARIOS |  |
| Si deseas profundizar en los temas revisados, te sugerimos consultar:   1. **Machine Learning Algorithms**   Bonaccorso, G. (2017). *Machine learning algorithms*. Packt Publishing Ltd.   1. **Data Science**   Verma, G. K., Soni, B., Bourennane, S., & Ramos, A. C. (2021). *Data Science*. Springer Singapore.   1. **Machine learning: the art and science of algorithms that make sense of data**   Flach, P. (2012). *Machine learning: the art and science of algorithms that make sense of data*. Cambridge university press. | *Formato APA* |

Es el proyecto de investigación que los alumnos comenzaron a desarrollar en la materia de Seminario de proyectos I

Es la referencia 1 de los Recursos complementarios

El detalle de esta parte es que no todos los proyectos de los alumnos tienen las mismas necesidades de preprocesamiento por lo cual estas referencias son un menú de las opciones que tienen para abordar su problemática.

Mismo caso que con la referencia anterior.