

SILABO

FACULTAD
ESCUELA PROFESIONAL
PROGRAMA DE ESTUDIOS

INGENIERIA ESTADISTICA E INFORMATICA
INGENIERIA ESTADISTICA E INFORMATICA
CARRERA PURA

I. INFORMACIÓN GENERAL

1.1 Identificación Académica

a) Curso	CIENCIA DE DATOS I
b) Código	EST324
c) Prerequisito	EST320 - APRENDIZAJE SUPERVISADO
d) Número de Horas	03h teóricas, 02h prácticas, Total 05 horas
e) Créditos	04
f) Número de Horas virtuales	02
g) Año y Semestre Académico	2025-II
h) Ciclo de Estudios	VIII
i) Duración	Del 25 de Agosto al 19 de Diciembre del 2025 (17 semanas)
j) Área Curricular	Estudios de especialidad
k) Características del Curso	Investigación, desarrollo e innovación

1.2 Docente

a) Apellidos y Nombres	ALEMAN GONZALES LEONID
b) Condición y Categoría	Docente Nombrado a TC, Categoría Asociado
c) Especialidad	Ingeniería Estadística e Informática, Ingeniería de Software

1.3 Ambiente donde se realizó el aprendizaje

- a) EI-202 y Laboratorio de Cómputo 02

II. SUMILLA

El curso de Ciencia de Datos I pertenece al área de especialidad, es de carácter teórico-práctico y tiene como propósito contribuir al desarrollo integral del estudiante, desarrollando metodologías esenciales para trabajar con datos de diversa característica en el desarrollo de análisis de tipo global basados en sistemas computacionales avanzados que permiten entender el comportamiento de la sociedad y resolver problemas complejos. Los contenidos fundamentales del curso incluyen: Gestión de datos digitales (Digital data), aprendizaje de técnicas learning avanzadas y aplicaciones. El curso está dividido en dos unidades:

- Unidad I : Introducción a la Ciencia de Datos, Metodologías y Recuperación de Información.
- Unidad II: Aprendizaje por técnicas learning avanzadas.

III. PERFIL DEL EGRESADO EN RELACIÓN AL CURSO

CE2. Construye y analiza modelos estadísticos y/o algoritmos computacionales para la toma de mejores decisiones en condiciones de incertidumbre, para el desarrollo de la región y el país, asumiendo una actitud científica y ética, utilizando técnicas estadísticas y/o informáticas referidas a la ciencia de datos.

IV. LOGRO DE APRENDIZAJE DEL CURSO

Construye modelos avanzados de aprendizaje y aplicación utilizando metodología de Ciencia de Datos, recuperación de información y aplicación de conocimientos de matemática, estadística, ciencia de la computación orientada a un dominio específico.

V. TRATAMIENTO DE UNIDADES DIDÁCTICAS

UNIDAD 1		Introducción a la Ciencia de Datos, Metodologías y Recuperación de Información
LOGROS DE APRENDIZAJE DE LA UNIDAD		
Aplica la perspectiva de ciencia de datos en la solución de problemas para el cual plantea los procesos en un proyecto utilizando técnicas digitales del manejo de información.		
TIEMPO DE DESARROLLO		Del 25 de Agosto al 27 de Octubre del 2025 (Total 45 horas)
HORAS DE ENSEÑANZA VIRTUAL/UNIDAD		02
SEMANAS	CRITERIOS DE DESEMPEÑO	CONOCIMIENTOS
Semana 1	Explica fundamentos de Ciencia de Datos, tecnologías, terminologías y el planteamiento de proyectos para la solución de problemas.	Fundamentos de Ciencia de Datos, Epistemología, procesos, proyectos, terminología y herramientas tecnológicas. de la Ciencia de Datos, artículo sobre ciencia de datos.
Semana 2	Analiza las principales metodologías en Ciencia de Datos, aplica la metodología CRISP-DM en escenarios reales y la adapta al contexto local.	Metodologías de Ciencia de datos KDD, OSEMN, TDSP, otros. Metodología CRISP-DM y uso de herramienta CRISP-DM.
Semana 3	Elabora el diseño conceptual de un proyecto de ciencia de datos contextualizado en una problemática real.	Proyectos de Ciencia de Datos, identificación de variables, objetivos, fuentes de datos, hipótesis, canvas de Ciencia de Datos y datasets abiertos.

Semana 4	Extrae temas clave desde documentos no estructurados usando técnicas de modelado de tópicos.	Comprensión del negocio (CRISP-DM) mediante minería de texto, algoritmos de LDA/NMF, evaluación de coherencia temática, herramientas spaCy, scikit-learn, pyLDAvis. Retroalimentación 1
Semana 5	Implementa un flujo de extracción de datos desde fuentes web, respetando restricciones legales y técnicas	Comprensión de los datos (CRISP-DM) con web scraping, HTML, DOM, XPath, Requests, parsers, APIs, Crawling, ética y robot.txt, requests, BeautifulSoup, Selenium, Scrapy, APIRest
Semana 6	Analiza y diagnostica conjuntos de datos con herramientas automatizadas	Comprensión de los datos (CRISP-DM) con exploración automatizada y detección de anomalías. Exploración, anomalías e informes automáticos. Algoritmos de Isolation Forest, LOF, clustering para detección, Proling de datos. Herramientas pandas-profiling (YData), PyCaret.
Semana 7	Genera y selecciona características relevantes usando técnicas automatizadas basadas en importancia e interpretabilidad.	Preparación de datos (CRISP-DM) con Feature Engineering. Feature generation, selection, interpretabilidad, Algoritmos SHAP, Boruta, Featuretools. Herramientas featuretools, shap, boruta_py.
Semana 8	Diseña y entrena redes convolucionales para tareas de clasificación de imágenes.	Modelado (CRISP-DM) con Redes Neuronales Convolucionales (CNN), Arquitectura CNN, entrenamiento e hiperparametrización. Algoritmos CNN clásica, Dropout, data augmentation. Herramientas PyTorch, TensorFlow, Keras, OpenCV.
Semana 9	Demostración de capacidades en evaluación.	Evaluación parcial. Presenta el diseño del proyecto con los avances en comprensión y preparación de datos. Retroalimentación.
PORCENTAJE DE AVANCE ACADÉMICO DE LA UNIDAD: 55%		

* El docente deberá programar “actividades de retroalimentación” los cuales se desarrollarán a lo largo de cada unidad didáctica con el fin de asegurar los adecuados aprendizajes de los estudiantes.

UNIDAD 2		Aprendizaje por técnicas learning avanzadas.
LOGROS DE APRENDIZAJE DE LA UNIDAD		
TIEMPO DE DESARROLLO		Del 27 de Octubre al 19 de Diciembre del 2025 (Total 40 horas)
HORAS DE ENSEÑANZA VIRTUAL/UNIDAD		00
SEMANAS	CRITERIOS DE DESEMPEÑO	CONOCIMIENTOS
Semana 10	Aplica CNNs a tareas reales de reconocimiento visual de objetos o patrones.	Modelado (CRISP-DM) de aplicación CNN al reconocimiento de objetos. Modelos entrenados, Fine-tuning, detección vs clasificación. Algoritmos VGG, ResNet, Yolo Vx. Herramientas torchvision.models, OpenCV, YOLOv5.
Semana 11	Entrena un modelo auto-supervisado para representación visual y aplica fine-tuning con pocas etiquetas.	Modelado (CRISP-DM) con aprendizaje auto-supervisado learning (etiquetado), simCLR, MoCo, BYOL. Algoritmos SimCLR, Triplet loss. Herramientas PyTorch Lightning, SimCLR, Kornia.
Semana 12	Implementa un flujo de entrenamiento basado en auto-supervisión textual con aplicación conversacional.	Modelado (CRISP-DM) con aprendizaje auto-supervisado aplicado a chatbots, pre-entrenamiento de lenguaje, BERT vs GPT, Transferencia y fine-tuning. Algoritmos Masked LM (BERT), Causal LM (GPT), Fine-tuning para clasificación o generación. Herramientas transformers, spaCy, OpenAI API, HuggingFace datasets.
Semana 13	Optimiza el umbral de decisión y analiza curvas para maximizar el rendimiento real del modelo.	Evaluación (CRISP-DM) orientada a optimización, ROC, PR Curve, AUC, FI, Recall, Umbral como hiperparámetro. Algoritmos ROC/PR Curve, Youden's J, Grid Search. Herramientas scikit-learn, Optuna, calibration_curve. Retroalimentación.

Semana 14	Construye y evalúa un sistema recomendador basado en contenido o colaborativo.	Aplicación Sistemas Recomendadores, filtrado colaborativo vs basado en contenido, métricas de evaluación: precision@k, recall@k. Algoritmos: KNN, Matriz Factorization (SVD, ALS), TF-IDF + Cosine similarity. Herramientas LightFM, surprise.
Semana 15	Implementa el modelo en un entorno accesible y funcional para el usuario.	Despliegue (CRISP-DM) Modelos Recomendadores, APIs REST, exportación de modelos, visualización interactiva. Modelos entrenados. Herramientas Flask, FastAPI, Streamlit, Docker.
Semana 16	Analiza implicancias éticas y sociales del uso de modelos predictivos, proponiendo soluciones.	Ciencia de Datos Responsable, sesgo algorítmico, equidad, privacidad, explainable AI. Algoritmos SHAP, LIME, Fairness-aware learning. Herramientas shap, aif360, Fairlearn, lime.
Semana 17	Evaluación: Presenta y defiende un proyecto completo de ciencia de datos aplicado, alineado a CRISP-DM.	Evaluación de conocimientos Presentación de proyecto final. Defensa del proyecto. Retroalimentación
PORCENTAJE DE AVANCE ACADÉMICO DE LA UNIDAD: 0%		

* El docente deberá programar “actividades de retroalimentación” los cuales se desarrollarán a lo largo de cada unidad didáctica con el fin de asegurar los adecuados aprendizajes de los estudiantes.

VI. ESTRATEGIAS METODOLÓGICAS

6.1 De Enseñanza

- Actividades síncronas:
- Actividades en video conferencias.
- Implementación de foros de discusión en línea
- Uso de sala de chat en línea.
- Actividades asíncronas:
- Revisión de foros de información sobre la temática
- Planteamiento y revisión de tareas
- Uso de materiales de lectura
- Proporcionar diapositivas con la temática y videos

6.2 De Aprendizaje

- Análisis de textos y artículos
- Revisión de guías y manuales
- Plataforma en la nube de programación Google COLAB, Jupyter Notebook.
- Debate, análisis y discriminación de ideas.
- Uso de software de lenguajes R y Python y plataformas de programación en Pc y la nube.
- Resolución de casos y exposición de propuestas de solución y análisis de soluciones
- Software IDEs, Pycharm, Jupyter, Spyder, Google Colaboratory
- Resolución de problemas y casos

6.3 De Investigación Formativa

- Revisión de recursos bibliográficos (libros, artículos y tesis) de forma estructurada.
- Preprocesamiento y transformación de datos.
- Presentación de informes y trabajos usando las normas y herramientas de publicación de artículos y reportes de investigación (estructura, partes, gráficos, citas, referencias, similaridad).

6.4 De Responsabilidad Social Universitaria

Analizar problemas de la sociedad a ser abordados, desarrollar la extracción de datos y proponer soluciones buscando aplicar los conocimientos del curso, gestionando un banco de datos y proyectos abiertos.

6.5 De Enseñanza Virtual

Uso de aula virtual, materiales digitales y recursos de aprendizaje de la nube, así como la administración y almacenamiento de recursos virtuales.

VII. MEDIOS Y MATERIALES DIDÁCTICOS

- Uso de plataformas web, máquinas virtuales y servicios web.
- Diapositivas, manuales y guías de aprendizaje digital.
- Materiales en video y pizarra en línea.
- Software IDE y framework para Python y R.

VIII. EVALUACIÓN DEL APRENDIZAJE

8.1 Logro de aprendizaje, evidencias de desempeño, ponderación, técnicas e instrumentos de evaluación.

UNIDAD	LOGROS DE APRENDIZAJE	EVIDENCIAS DESEMPEÑO: De acción, objeto o producto (%)	PONDERACIÓN (Obligatorio en base a 100%)	TÉCNICAS	INSTRUMENTOS
1	Aplica la perspectiva de ciencia de datos en la solución de problemas para el cual plantea los procesos en un proyecto utilizando técnicas digitales del manejo de información.	Evidencia 1: Informe Defensa Evidencia 2: Examen t/p - Resolución de casos - Implementación	50%	- Revisión de informe - Observación - Examen - Análisis de resultados	- Lista de cotejo - Registro - Rúbrica - Lista de cotejo - Informe
2		- Lista de cotejo - Registro - Rúbrica - Lista de cotejo - Informe	50%	- Revisión de informe - Observación - Examen - Análisis de resultados	- Revisión de informe - Observación - Examen - Análisis de resultados

8.2 Evidencias de aprendizaje del semestre académico.

LOGRO DE APRENDIZAJE DEL CURSO	EVIDENCIAS: ACCIÓN, OBJETO o PRODUCTO	FECHA DE PRESENTACIÓN
Construye modelos avanzados de aprendizaje y aplicación utilizando metodología de Ciencia de Datos, recuperación de información y aplicación de conocimientos de matemática, estadística, ciencia de la computación orientada a un dominio específico.	Desarrollo de modelos de aprendizaje con técnicas avanzadas a problemas complejos con recuperación de información y metodologías de ciencia de datos.	09 al 12 de diciembre 2025.

• Evidencias que deben ser socializados con la comunidad como parte de la responsabilidad social universitaria. Por la emergencia sanitaria podría optarse por la difusión vía página web y redes sociales.

8.3 Calificación

La fórmula para la obtención del promedio final del curso es la siguiente:

$$\text{Promedio Final} = (50\%)IUPP + (50\%)IIUPP$$

Donde:

IUPP : Primero unidad promedio parcial

IIUPP : Segundo unidad promedio parcial

La fórmula para la obtención del promedio final del curso es la siguiente:

$$\text{Promedio Final} = (50\%)IUPP + (50\%)IIUPP$$

Donde:

IUPP : Primero unidad promedio parcial

IIUPP : Segundo unidad promedio parcial

$$\text{Promedio parcial de la unidad} = (50\% (LA1) + 50\% (LA2)) / (100\%)$$

LA1 = Logro de aprendizaje 1.

LA2 = Logro de aprendizaje 2.

La fórmula para la obtención del promedio final del curso es la siguiente:

$$\text{Promedio final} = (IUPP + IIUPP) / 2$$

IX. FUENTES DE INFORMACIÓN

9.1 Bibliográficas

Básica

- Provost, F., & Fawcett, T. (2013). Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking. O'Reilly Media.
- VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.
- Wickham, H., & Golemund, G. (2016). R for Data Science: Import, Tidy, Transform, Visualize, and Model Data. O'Reilly Media.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.
- Géron, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd ed.). O'Reilly Media.
- Chollet, F. (2018). Deep Learning with Python. Manning Publications.
- Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction (2nd ed.). MIT Press.
- Silver, D. (2015). Reinforcement Learning Course. University College London. (Nota: curso disponible en línea)
- O'Neil, C. (2016). Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing Group.
- D'Ignazio, C., & Klein, L. F. (2020). Data Feminism. MIT Press.
- Mittelstadt, B. D., et al. (2016). The Ethics of Algorithms: Mapping the Debate. Big Data & Society, 3(2), 1–21.

Complementarias

- Floridi, L. (2013). The Ethics of Information. Oxford University Press.
- Eubanks, V. (2018). Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor. St. Martin's Press.
- Weng, L. (2018). Policy Gradient Algorithms. ArXiv preprint.
- Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement Learning: A Survey. Journal of Artificial Intelligence Research, 4, 237–285.
- Norvig, P., & Russell, S. (2021). Artificial Intelligence: A Modern Approach (4th ed.). Pearson.
- Nielsen, M. A. (2015). Neural Networks and Deep Learning. Determination Press.
- Brownlee, J. (2019). Deep Learning for Time Series Forecasting. Machine Learning Mastery.
- Kelleher, J. D., & Tierney, B. (2018). Data Science. The MIT Press.
- Warden, P. (2011). Big Data Glossary. O'Reilly Media.

Electrónicas

- Harvard University. (n.d.). CS109A: Introduction to Data Science. Retrieved from <https://cs109.github.io>
- Kaggle. (n.d.). Learn Data Science. <https://www.kaggle.com/learn>
- Google. (n.d.). Google Dataset Search. <https://datasetsearch.research.google.com/>
- Stanford University. (n.d.). CS231n: Convolutional Neural Networks for Visual Recognition. <https://cs231n.stanford.edu>
- DeepLizard. (n.d.). Deep Learning Playlists. YouTube. <https://www.youtube.com/c/deeplizard>
- Papers with Code. (n.d.). State-of-the-art Deep Learning Papers. <https://paperswithcode.com/>
- David Silver. (n.d.). Reinforcement Learning Course. DeepMind. <https://www.davidsilver.uk/teaching>
- Berkeley AI. (n.d.). CS 285: Deep Reinforcement Learning. <https://rail.eecs.berkeley.edu/deeprlcourse/>
- GitHub - RL Examples. <https://github.com/dennybritz/reinforcement-learning>

Producción intelectual del docente relacionado con el curso

- Guías de programación con Python para Machine Learning (2019).

Puno, Diciembre del 2025