

UNIVERSIDAD MAYOR, REAL Y PONTIFICIA DE SAN FRANCISCO XAVIER DE CHUQUISACA

FACULTAD DE TECNOLOGÍA



Primer Parcial

Carrera : Ingeniería en Ciencias de la Computación **CU:**11-453

Universitario : Lujan Renteria David Fernando

Sucre – Bolivia

2024

- **Dataset:**
[Forest Cover Type Dataset \(kaggle.com\)](#)
- **Git(Contiene todo en un cuadernillo y todo no necesario esta incluido en el Dataset):**
[SIS-420/Parciales/Primer_Parcial at main · lujan-99/SIS-420 \(github.com\)](#)
- **Drive(Estan todos los bloques separados ademas de los dataset utilizados):**
<https://drive.google.com/drive/folders/1OWEvKXKdbXiNkwKP8o78KW54ajsLShiZ?usp=sharing>

1. Análisis del Dataset

- 1.1 Análisis de los Valores del Dataset
- 1.2 Transformación de Variables Categóricas
- 1.3 Generación de Nuevas Columnas
- 1.4 Limpieza del Dataset
- 1.5 Combinar Datasets
- 1.6 Preparación del Dataset

2. Generación de Datos Sintéticos

- 2.1 Distribuciones Basadas en los Datos Originales
- 2.2 Generación de Datos Categóricos
- 2.3 Atributos Calculados

3. Verificación entre el Dataset Sintético y el Real

4. Limpieza y Preparación del Dataset

- 4.1 Combinar los Datasets
- 4.2 Transformación de Variables Categóricas y Manejo de Valores Faltantes
- 4.3 Normalización de Variables Numéricas
- 4.4 Guardar el Dataset Preparado

5. Implementación de Regresión Logística para el Dataset

- 5.1 Función Sigmoide

- 5.2 Función de Costo con Regularización L2
- 5.3 Descenso por Gradiente con Regularización L2

6. One-vs-All

- 6.1 Implementación de One-vs-All
- 6.2 Predicción con One-vs-All
- 6.3 Cargar y Preparar los Datos
- 6.4 Normalizar las Características
- 6.5 Dividir los Datos en Entrenamiento y Prueba