

Esta es una copia de lo que está en el README

Predicción de la Producción Industrial Mensual por Sector en Tierra del Fuego

Materia: Aprendizaje Automático

Carrera: Tecnicatura Superior en Ciencia de Datos e Inteligencia Artificial

Institución: Politécnico Malvinas Argentinas

Alumno: Luciano Luna

Año: 2025

Descripción del Proyecto

Este proyecto propone desarrollar un modelo de regresión supervisada capaz de predecir la producción industrial mensual en la provincia de Tierra del Fuego AELAS, desagregada por sector. Para ello, se integraron tres fuentes de datos oficiales sobre empleo, establecimientos y producción industrial. Se busca identificar patrones entre las variables estructurales (empleo y establecimientos) y el comportamiento productivo de sectores clave como el electrónico, textil, plástico, entre otros.

Objetivo

Objetivo General:

Construir un modelo de regresión supervisada que estime la producción mensual por sector industrial en función de:

- La cantidad de empleados por sector industrial.
- El número de establecimientos activos por sector.
- Variables temporales: mes y año.

Objetivos Específicos:

- Integrar y limpiar datasets históricos industriales.
- Unificar los datos en un único dataset estructurado por sector, mes y año.
- Establecer una base de entrenamiento para aplicar modelos de predicción.

Preguntas de Investigación

- ¿Es posible predecir el volumen de producción mensual en un sector a partir de las demás variables?
- ¿Cuáles son los sectores más sensibles a los cambios en empleo y cantidad de establecimientos?
- ¿Existen patrones estacionales o tendencias particulares en algunos sectores?

Contexto y Relevancia

Tierra del Fuego posee una estructura productiva altamente concentrada en la industria manufacturera. La posibilidad de anticipar caídas o aumentos en la producción por sector representa una herramienta de gran utilidad para la planificación económica regional, el diseño de políticas públicas, y la toma de decisiones en sectores industriales estratégicos. Este proyecto transforma datos administrativos en información valiosa para la gestión y el análisis prospectivo.

Modelo Propuesto

Este proyecto se enmarca en una tarea de aprendizaje supervisado de tipo regresión, cuyo objetivo es predecir la cantidad mensual de producción industrial por sector industrial en Tierra del Fuego.

Las variables independientes disponibles (empleados, establecimientos, año, mes) permiten entrenar modelos de regresión con el objetivo de estimar la variable dependiente producción.

Modelos sugeridos a utilizar:

- Regresión Lineal Múltiple
- Árboles de Decisión para regresión
- Random Forest Regressor
- Regresión Ridge (regularizada)

Estas técnicas permitirán evaluar tanto relaciones lineales como no lineales, y considerar posibles interacciones entre variables temporales y estructurales del sector industrial.

Descripción de los Datasets

Dataset	Fuente	Años	Variables principales
Producción Industrial por sector	Instituto Provincial de Estadísticas y Censos (TDF AeIAS)	2013–2025	Año, mes, sector, volumen de producción
Empleo en establecimientos industriales	Instituto Provincial de Estadísticas y Censos (TDF AeIAS)	2001–2025	Año, mes, sector, empleados
Cantidad de establecimientos industriales	Instituto Provincial de Estadísticas y Censos (TDF AeIAS)	2001–2025	Año, mes, sector, establecimientos

Origen y Descarga de los Datos

Los datos fueron descargados desde el sitio oficial del Instituto Provincial de Estadísticas y Censos (IPEC) de Tierra del Fuego el día 13 de mayo de 2025.

Los archivos originales se encuentran en la carpeta data/raw del repositorio:

- 14_3_01_Personal_industria_rama-1.xlsx → Personal ocupado.
- 14_3_02_Establecimientos_industriales_rama-1.xlsx → Cantidad de establecimientos.
- 14_3_03_Produccion_Industrial-1-1.xlsx → Producción mensual.

Los tres datasets fueron preprocesados para:

- Normalizar los nombres de sectores y meses.
- Unificar las estructuras.
- Resolver inconsistencias y valores nulos o atípicos.

Estructura del Dataset Final

El dataset consolidado resultante contiene las siguientes variables:

Variable	Tipo de Dato	Descripción
anio	int64	Año del registro (2001 a 2025)
mes	object	Mes del registro (en minúsculas, e.g., "enero")
sector	object	Sector industrial (e.g., "Textiles" , "Pesqueras")
Produccion	float64	Producción total mensual del sector (en unidades, según sector)
empleados	int32	Cantidad de empleados del sector ese mes
establecimientos	int32	Cantidad de establecimientos industriales activos ese mes

El dataset final tiene por el momento 1746 filas y 6 columnas: anio, mes, sector, Produccion, empleados, establecimientos. Los tipos de datos fueron cuidadosamente transformados para garantizar compatibilidad con modelos de regresión y visualizaciones.

Estructura del Repositorio - 2DA ENTREGA

└─ data/ | └─ raw/ <- Archivos Excel originales | └─ processed/ <- CSVs
intermedios y finales | └─ notebooks/ | └─ 01_exploracion_y_union_datasets.ipynb
<- Notebook principal de integración y limpieza | └─ README.md <- Este archivo

Pasos realizados

1. Carga de datos crudos: se importaron 3 archivos .xlsx correspondientes a producción, empleo y establecimientos industriales por sector.
2. Limpieza y transformación:
 - Se identificaron estructuras no tabulares (por ejemplo, subtítulos, múltiples hojas y encabezados intermedios).
 - Se normalizaron columnas como mes y sector para asegurar la consistencia.
 - Se eliminaron columnas innecesarias y registros vacíos.
 - Se imputaron valores faltantes específicos en sectores como Pesqueras, aplicando interpolación manual basada en el promedio entre el mes anterior y el posterior.
3. Unificación de datasets:
 - Se realizó un merge progresivo de los tres datasets procesados usando como claves las variables anio, mes y sector.
 - Se mantuvieron los valores faltantes (NaN) en las columnas de producción en años donde no había datos disponibles.

4. Exportación de datos:

- Cada dataframe intermedio fue exportado a la carpeta data/processed/.
- El dataset unificado se utilizó como base para la modelización en futuras entregas.