

Grado en

**Business Data Analytics**

Informe final

Reto: XXX [número y nombre del reto]

Curso: 2º 2025-2026

Equipo: [nombre del equipo]

URL del repositorio en GitHub:

Ruta de la carpeta de Google Drive que contiene el proyecto:

(Ej. AZUL\_CLARO\R5\_Proyecto\_AZUL\_CLARO)

Autores:

* Erik Arbaiza
* Jon Basarte
* Adriana Díez
* Iraia Laraudogoitia
* Aketx Tejedor
* Gabriel Velasco

[Nombre del fichero] 25-26\_2\_R5\_AzulClaro

# Índice

## Introducción

*Explicar los apartados que contiene este informe, si ha habido alguna limitación, algo que sea importante saber para entenderlo mejor. Contenido que sirve para “preparar” a quien lo vaya a leer.*

## Identificar la problemática

### Sobre el cliente

Laboral Kutxa es una cooperativa de crédito vasca perteneciente al Grupo Mondragón, referente internacional en el ámbito de la economía cooperativa. Desde su creación, su objetivo principal ha sido ofrecer a las cooperativas industriales servicios sociales, económicos y empresariales.

Su misión consiste en atender las necesidades financieras de socios y clientes, apostando por el liderazgo en calidad y servicio a través de un asesoramiento cercano y personalizado. Los más de 2.200 socios trabajadores que forman parte de la organización están comprometidos con una forma diferente de hacer banca, en la que la ética, la cooperación y la sostenibilidad constituyen pilares fundamentales.

El compromiso con el medio ambiente es otro de los ejes estratégicos de Laboral Kutxa. Por ello, cuentan con un Sistema de Gestión Ambiental y desarrollan numerosas iniciativas para que sus actividades resulten más sostenibles. De hecho, son pioneros en el sector financiero en implantar este tipo de sistema en los edificios de su sede central.

### Sobre la problemática

La problemática central del proyecto consiste en generar predicciones fiables del **PIB** (Producto Interno Bruto) y del **IPC** (Índice de Precios al Consumidor) de Australia para el último trimestre del año 2022. Ambos indicadores son fundamentales para **evaluar la situación económica** y **la inflación del país**, información que resulta clave para que el cliente tome **decisiones estratégicas** sobre su posible **expansión en el mercado australiano**.

Es por ello, que se analizarán los datos proporcionados por la empresa, que incluyen los indicadores mencionados junto con diversas variables **macroeconómicas** relevantes. Posteriormente, se compararán las predicciones obtenidas con estimaciones realizadas por **expertos**, con el fin de identificar la fiabilidad de los modelos y proporcionar un análisis sólido sobre las perspectivas económicas de Australia.

### Objetivos

Con el fin de apoyar a Laboral Kutxa en su posible entrada al mercado australiano, este proyecto busca predecir la evolución del PIB y del IPC de Australia para el último trimestre de 2022, utilizando modelos desarrollados en Fundamentos Matemáticos y Data Mining. Estas predicciones permitirán analizar la situación económica del país y proporcionar información relevante sobre tendencias de crecimiento o desaceleración.

A partir de los resultados obtenidos, se elaborarán recomendaciones estratégicas que ayuden al cliente a evaluar la viabilidad y las condiciones del mercado australiano, considerando posibles acciones para su expansión. El objetivo es ofrecer un análisis claro que sirva como base para la toma de decisiones estratégicas.

Además, se compararán las predicciones propias con estimaciones realizadas por expertos, como autoridades oficiales y entidades financieras, y se contrastarán ambas con los valores reales de PIB e IPC publicados para el último trimestre de 2022. Esta comparación permitirá evaluar la fiabilidad de los modelos utilizados y fortalecer las conclusiones y recomendaciones del proyecto.

## Recoger y almacenar los datos

### Fuentes de datos utilizadas

*Breve descripción de las bases de datos proporcionadas*

### Procesamiento de los datos

Para poder completar los objetivos propuestos, se comenzó filtrando y limpiando los datos, considerando únicamente el país asignado, Australia. Se contaba con bases de datos que incluían variables exógenas y otra base con las variables principales, PIB e IPC. Los datos fueron filtrados y consolidados en un único data frame mediante un left\_join, asegurando que todas las variables estuvieran alineadas temporalmente por país, código, continente, año y mes. Durante esta revisión inicial se identificaron **valores faltantes** en algunas variables mensuales clave, que fueron imputados manualmente utilizando fuentes confiables como **Trading Economics y FRED,** garantizando la continuidad de las series hasta septiembre de 2022.

Posteriormente, se procedió a transformar todas las variables a **series temporales trimestrales.** Las variables exógenas, originalmente mensuales, se agregaron de manera diferenciada según su naturaleza: la tasa de desempleo se calculó como **promedio trimestral**, mientras que la oferta monetaria y el índice bursátil tomaron el **valor correspondiente al último mes de cada trimestre**. El PIB y el IPC ya estaban disponibles en frecuencia trimestral, por lo que solo fue necesario organizar sus valores y eliminar los NAs iniciales. Esta homogeneización permitió disponer de un conjunto consistente de series temporales, fundamental para aplicar modelos predictivos y realizar comparaciones coherentes entre todas las variables económicas.

Con el fin de explorar la evolución de los datos y asegurar su calidad, se generaron gráficos de las series mensuales y trimestrales para cada variable, identificando patrones, estacionalidades y valores atípicos. Los **outliers** fueron detectados con tsoutliers() y visualizados sobre la serie original mediante puntos que señalaban su ubicación. Posteriormente, se eliminaron utilizando tsclean(), reemplazándolos por interpolaciones robustas. La mayoría de los valores atípicos **eran puntuales y dispersos**, por lo que la limpieza no afectó la estructura general de las series, garantizando que los análisis posteriores y los modelos predictivos del PIB e IPC se realizaran sobre datos consistentes y representativos de la evolución económica australiana.

Este procesamiento de datos permitió disponer de un **conjunto sólido y uniforme,** asegurando que los modelos predictivos generen resultados **fiables** y reflejen patrones reales de la economía australiana, proporcionando a Laboral Kutxa información precisa para evaluar oportunidades de expansión y comparar sus predicciones con estimaciones oficiales y de expertos.

## Analizar y modelar los datos

### Análisis descriptivos

### Modelar los datos y Visualizar los resultados obtenidos

Para llevar acabo las predicciones, se dividierons las series temporales del PIB y IPC en conjuntos de entrenamiento y prueba. El entrenamiento incluyiendo el primer tirmestre de 1998 hasta cuarto trimentes de 2020, mientras que las series de prueba incluyeron desde el primer trismeste de 2021 hasta el segundo de 202.

Una vez se garantiza la estacionariedad de las series, se procedió a ajustar diferentes modelos de series temporales, AutoARIMA, ARIMA manual y SARIMA. La selección de los parámetros para los modelos ARIMA y SARIMA manuales, la selección de los parámetros “p, d, q” y estacionales “P,D,Q”, se baso en el análisis de las funciónes de autocorrelación (ACF) y autocorrelación parcial (PACF) de las series diferenciadas. La validación de los modelso se realizo evaluando los residuos mediante el test de Ljung -box y la inspección grafica de residuos con la función checkresiduals. En general los residuos de los moleos del IPC y de PIB se comportaron como ruido blanco, con la excepción del modelo AutoARIMA del PIB, que presento autocorrelación significativa (p-value = 2.153e – 05). Los gráficos de la validación de reisudos y los detalle del tecto de LJUNG se presentan en el ANEXO 1 .

Las predicciones se realizaron sobre el conjunto de prueba utilizando la función forecast(). Dado que las series habían sido tranformadas o diferenciadas, los pronósticos fueron revertidos a las unidades originales antes de compararlos con los datos observados. En el caso de IPC, que fue diferenciada una sola vez, se utilizo diffinv() para revertir la diferencia simple. Para el PIB, que requierio estabilización de la varianza y diferencias dobles y un lag de cuatro trimestres, se aplico la inversión de ambas diferencias (primero la no estacional y luego la estacional con lag =4) y se deshizo la transformación logarítmica. De esta forma, los pronósticos se expresaron en las unidades originales, permitiendo una comapracion directa con los valores reales observados.

A continuación se comaprarons las métricas de cada modelo para cad ana de las variables.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| MODELOS IPC | | | | |
| **Modelo** | **ME** | **RMSE** | **MAE** | **MAPE** |
| **AutoARIMA** | 0.2302543 | 0.7874101 | 0.5962474 | 0.4993937 |
| **ARIMA** | 0.2882195 | 0.8083731 | 0.5982361 | 0.5004915 |
| **SARIMA** | 0.2904238 | 0.8084165 | 0.5975839 | 0.4999148 |

Se continuo con las predicción para los conjutos de prueba con al función forecast con un h de la longitud de test de ipc y test pib. Se prodecio haciendo la reversion ya que esos valores del forecast no estánban en la mismas unidades que las de la serie original, es por ello que pa el caso del ipc la serie pa volverla estacioanria se revirtió una sola ves por lo que mediante diffinv se consiguió revertir consigueindo así los valores reales. En el caso del pib la serie necesito ser la vairnza estabilizada y aplicarle dos diferencias con un log= 4 por que lo que la reversion….. una vez obtenidos todos los forecast de cada modelo revertidos en sus unidades originales se porcedio a comaprar el mejor modelo comparando sus accrucay

### Conclusiones

*En un par de párrafos o tres resumir qué se concluye de todo el análisis y modelos de los datos.*

## Transformar los negocios

## Implicaciones legales y éticas

En un entorno donde **cada vez más decisiones se toman mediante el uso de algoritmos y modelos predictivos**, los **aspectos éticos adquieren una relevancia fundamental**.

## Bibliografía

## Apéndice I

*Desarrollo matemático. Se os darán instrucciones concretas en las clases de la materia “Fundamentos matemáticos y estadísticos”*

## Apéndice II

*Big Data. Breve de explicación de los pasos realizados para la creación de la máquina virtual.*

ANEXO DATA MINING