

Proyecto Catch

Yeiver Sanabria

2025-10-29

1. Introducción

El presente proyecto tiene como objetivo analizar y pronosticar la empleabilidad en la industria manufacturera mexicana (Sector 31-33), utilizando datos históricos del Instituto Nacional de Estadística y Geografía (INEGI). Mediante un enfoque de análisis de series de tiempo, se busca comprender el comportamiento, las tendencias y los componentes estacionales del empleo en el sector, con el fin de anticipar períodos de alta contratación y ofrecer a las compañías de reclutamiento de talento humano una herramienta estratégica para planificar recursos y fortalecer su posicionamiento competitivo.

La industria manufacturera representa uno de los pilares económicos del país, aportando aproximadamente el 20 % del Producto Interno Bruto (PIB) nacional (Banco Mundial, 2024). En 2023, este sector alcanzó un valor de producción cercano a 5.2 billones de pesos mexicanos, consolidándose como uno de los rubros de mayor crecimiento y relevancia en la economía nacional (Instituto Nacional de Estadística y Geografía [INEGI], 2024). Este dinamismo hace del sector un terreno clave para la planificación de talento, ya que la capacidad de anticipar los ciclos de contratación puede traducirse en ventajas operativas y comerciales significativas.

De acuerdo con estimaciones del mercado, el costo promedio de reclutamiento en México oscila entre el 15 % y el 25 % del salario anual del candidato colocado, pudiendo alcanzar hasta el 30 % en perfiles especializados (Agencia de Reclutamiento, 2025; Del Moral, 2025). En este contexto, una firma especializada que logre posicionarse como socio estratégico del sector manufacturero podría generar ingresos superiores a 30 millones de pesos anuales, considerando un volumen estimado de 500 colocaciones con un salario promedio de 300 000 MXN por empleado. Para sustentar esta visión, se aplicaron tres modelos de pronóstico —Holt-Winters, ARIMA y Bayesian Structural Time Series (BSTS)—, logrando estimar la empleabilidad del sector hasta el año 2026, y consolidando una propuesta que combina análisis científico con valor comercial tangible.

2.Metodologia

2.1 Recoleccion de informacion

El insumo inicial para este proyecto, fueron datos tabulares recopilados directamente del INEGI, desde su pagina oficial, atraves de los servicios para desarrolladores, desde los cuales se exporto en formato json, una serie temporal con indicadores asociados a la empleabilidad por año en esta industria, para mas informacion, consulte el siguiente enlace: https://www.inegi.org.mx/servicios/api_indicadores.html#idMetodoIndicadoresInegi.

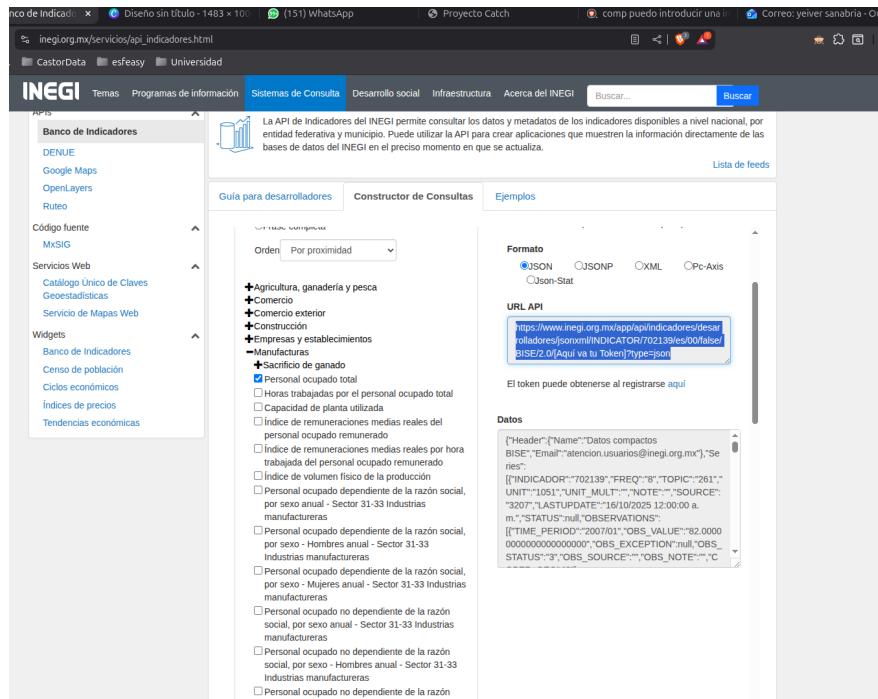


Figure 1: API INEGI

Ejecutando el siguiente curl tambien se pueden obtener los datos mencionados:

```
curl -sS 'https://www.inegi.org.mx/app/api/indicadores/desarrolladores/jsonxml/INDICATOR/702139/es/00/false/BISE/2.0/7354251f-cd53-e0e6-22f2-6f859ac860c2?type=json'  
-H 'Accept: application/json'
```

2.2 Transformacion

Con la informacion antes mencionada, se realizó una transformación de los datos. Inicialmente, con un formateador gratuito de JSON, se adaptó el formato de los datos y se construyó una tabla en formato CSV. Luego, tomando el total de personas empleadas para el año 2018, se transformó la data, la cual viene con unidad de medida: UNIT=1051, que según el diccionario

de datos del INEGI corresponde a índice base 2018. Por lo tanto, para tener toda la serie temporal en personas empleadas fue necesario hacer el siguiente cálculo:

$$\text{personas_mes} = 6,493,020 \times (\text{índice_mes} / 100)$$

Adicionalmente, se removieron las columnas que no eran de interés, dejando un archivo en formato CSV con una columna asociada a la fecha de recolección y una columna de personas, con el total de personas empleadas en la industria manufacturera, con información desde 2007 hasta 2025.

2.3 Análisis temporal

Para el análisis de la información recolectada, se utilizó RStudio. Inicialmente, se cargó el archivo CSV construido en las secciones 2.1 y 2.2.

```
data <- read.csv('/home/yeiver/Documentos/universidad/proyectoCatch/personas_mensuales_d  
attach(data)
```

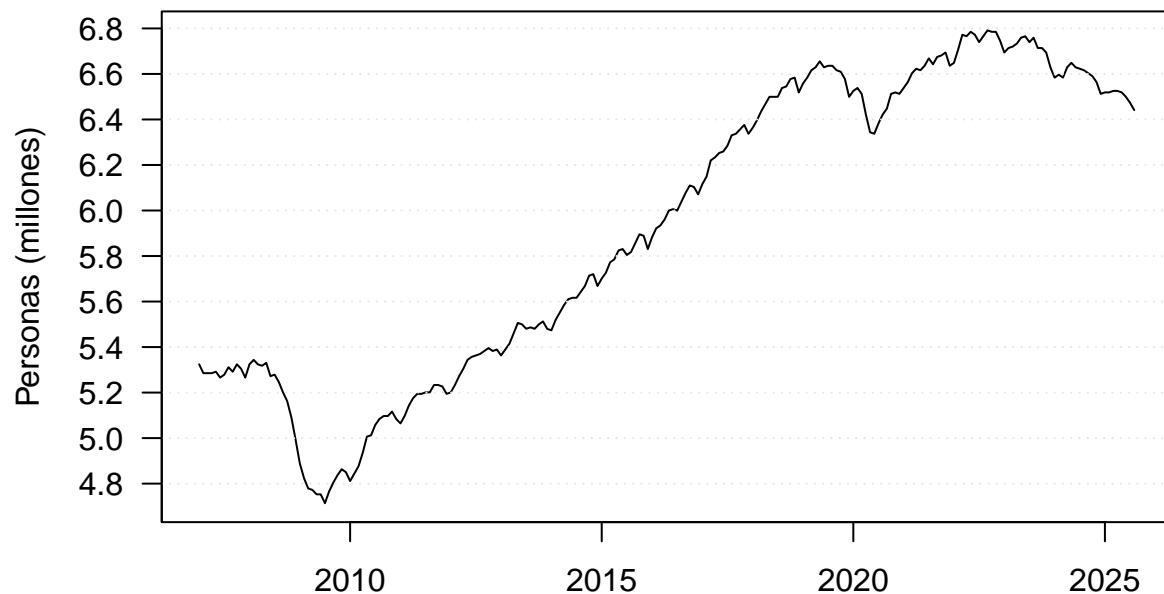
Luego, se realizó la transformación necesaria a la columna fecha, para que fuera identificada como tipo Date

```
data <- data %>%  
  mutate(`fecha` = ym(`fecha`))
```

Con el dataframe con los formatos adecuados, instanciamos la serie de tiempo.

```
manufactura_ts <- ts(data$personas, start = c(2007, 1), freq = 12)
```

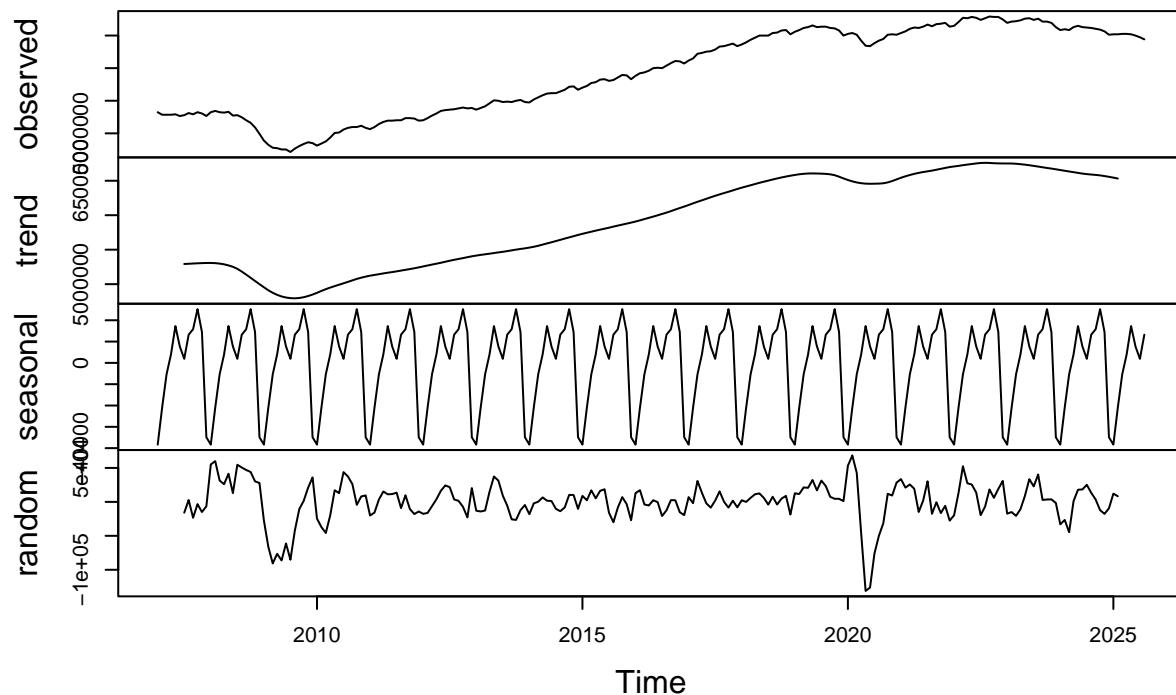
Cantidad de personas empleadas en la Industria Manufacturera en México (2007–2025)



Fuente: INEGI – Banco de Indicadores (BISE).

Como se puede observar, la serie presenta una tendencia creciente, aunque no constante. Esto refleja el claro crecimiento de la industria en el país y coincide con los datos que reporta el INEGI.

Decomposition of additive time series

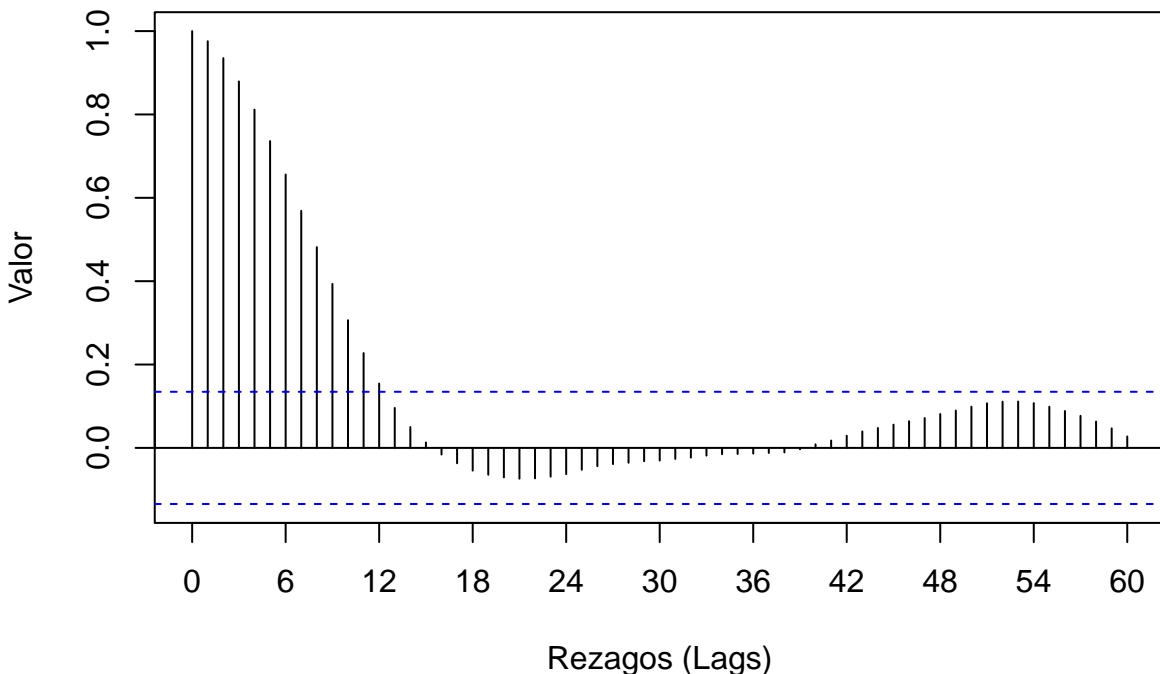


Un análisis más detallado de la serie de tiempo expone factores de gran relevancia. Como se mencionó anteriormente, parece haber una clara tendencia creciente en los últimos años, pero dicha tendencia presenta comportamientos depresivos en los últimos períodos. Adicionalmente, se puede concluir que hay un claro factor estacional o cíclico, lo que indica que la cantidad de personas empleadas en esta industria tiene una frecuencia definida. Esto puede indicar en qué meses aumenta la demanda de personal, información de gran utilidad que permite cuantificar lo que intuitivamente se conoce: que hay meses de mayor contratación. Teniendo en cuenta esta información, se pueden tomar decisiones estratégicas y ofrecer el personal adecuado en el momento oportuno para fortalecer el trabajo en la industria manufacturera.

```
diff_12 <- diff(manufactura_ts, lag = 12)

ac <- acf(diff_12, lag.max = 60, plot = FALSE)
```

Diferencia interanual del empleo manufacturero en Mexico



En la gráfica de correlaciones con lag 12 se puede identificar la influencia que tienen los períodos anteriores sobre los actuales. Esto otorga información relevante, ya que se puede afirmar con un 95 % de confianza que los resultados estadísticos de las correlaciones del año anterior tienen una influencia notoria sobre el período actual. Por lo tanto, conocer la empleabilidad en el año en curso nos muestra cómo se comportará el sector en el período siguiente. Se aproximará la serie temporal utilizando el modelo de Holt-Winters para determinar los valores estacionales —es decir, la influencia que tiene cada mes sobre la contratación— y hacer un

pronóstico del año siguiente. En consecuencia, se podrá determinar cuál será la demanda de personal en esta industria por cada mes del año 2026. Esta información nos brinda una visión de cómo se comporta el sector, permitiéndonos estar un paso adelante de competidores que también buscan este talento humano, incluso dentro del mismo sector, capacitando al personal y advirtiendo a las compañías sobre posibles fugas del recurso humano en los meses de influencia negativa.

3. Implementación de modelos

Para realizar una comparación adecuada entre modelos, la serie se dividió en dos: entrenamiento y prueba.

```
train <- window(manufactura_ts, end = c(2024, 12))
test  <- window(manufactura_ts, start = c(2025, 1), end = c(2025, 8))
```

Utilizando el año 2025, como periodo de prueba, para poder medir y comparar los modelos propuestos a continuacion.

3.1 Modelo de Holt-Winters

El modelo de Holt-Winters es un método de suavizamiento exponencial usado para pronosticar series de tiempo que muestran tendencia y estacionalidad. Combina tres componentes: nivel, tendencia y estacionalidad, ajustándolos de forma dinámica para hacer predicciones más precisas a corto plazo (Winters, 1960).

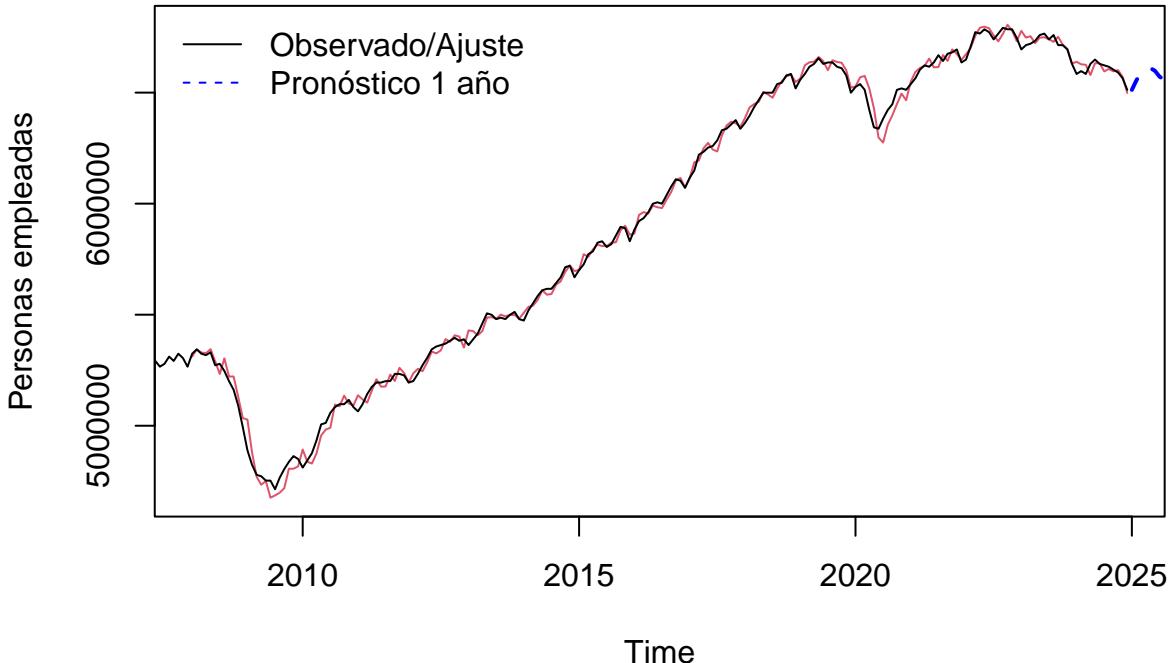
```
train_hw <- HoltWinters (train, seasonal = "mult")
train_hw$coef ; train_hw$SS

##           a             b             s1            s2            s3
## 6.584984e+06 -4.939883e+03 9.897090e-01 9.971091e-01 1.002613e+00
##           s4             s5             s6             s7             s8
## 1.005258e+00 1.007687e+00 1.006571e+00 1.003301e+00 1.001963e+00
##           s9             s10            s11            s12
## 9.998123e-01 1.000030e+00 9.978180e-01 9.889924e-01
## [1] 223550907935
```

En la tabla de pronóstico se observó cómo cada valor s_i corresponde al coeficiente estacional, que explica cómo cada mes altera el promedio de empleados en el sector manufacturero. Se concluye, con la información recolectada, que la tendencia creciente de la cantidad de empleados efectivamente se está reduciendo, según el parámetro b , que representa la tendencia en el último periodo observado.

```
## [1] 6550134140
```

Holt–Winters filtering



```
print(fc_hw)
```

```
##           Jan      Feb      Mar      Apr      May      Jun      Jul      Aug
## 2025 6512329 6556096 6587329 6599744 6610714 6598419 6572030 6558317
```

Como se puede observar en la grafica roja, el modelo se ajusta muy bien a los datos observados, reconoce tanto la tendencia creciente, como el factor estacional de la misma y se ajusta adecuadamente; el MSE nos da un valor razonable, en el contexto de trabajo, por lo que podemos afirmar que el modelo es funcional. Respecto al pronostico se analizan los valores mensuales generados por el modelo, y una vez graficados en la serie original parecen fluir con normalidad sobre el patron observado, por lo que se pueden utilizar como fuente aceptable de pronostico.

3.1 Modelo ARIMA

Una metodología más convencional para abordar problemas de series de tiempo es el modelo ARIMA. El modelo ARIMA (AutoRegressive Integrated Moving Average) es un método estadístico usado para pronosticar series de tiempo al combinar tres componentes: autorregresión (AR), diferenciación (I) y media móvil (MA), permitiendo capturar patrones de dependencia y tendencia en los datos (Box & Jenkins, 1970).

```

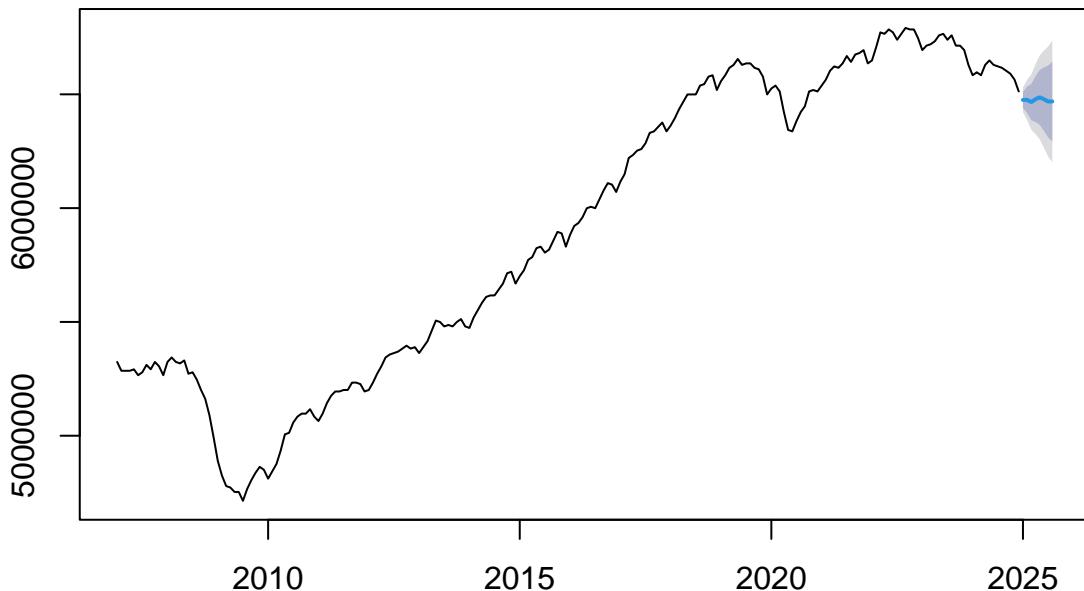
fit_arima <- auto.arima(train, seasonal = TRUE, stepwise = FALSE, approximation = FALSE)
summary(fit_arima)

## Series: train
## ARIMA(1,1,1)(2,0,0)[12]
##
## Coefficients:
##             ar1      ma1     sar1     sar2
##             0.7775 -0.4654  0.3306  0.2148
## s.e.    0.0859   0.1222  0.0682  0.0721
##
## sigma^2 = 751337856: log likelihood = -2501.94
## AIC=5013.88   AICc=5014.17   BIC=5030.74
##
## Training set error measures:
##               ME      RMSE      MAE      MPE      MAPE      MASE
## Training set 914.551 27091.43 21599.83 0.01855429 0.3726386 0.1241777
##               ACF1
## Training set 0.004245438

fc_arima <- forecast(fit_arima, h = h)
plot(fc_arima, main="Pronostico empleabilidad sector manufacturero Mexico 2025 ARIMA")

```

Pronostico empleabilidad sector manufacturero Mexico 2025 ARIMA



```

mse_arima <- mean((as.numeric(test) - as.numeric(fc_arima$mean))^2, na.rm = TRUE)
print(mse_arima)

## [1] 1472929647

```

El análisis de los coeficientes del modelo ARIMA indica que la dinámica del número de personas empleadas en la industria se caracteriza por una fuerte dependencia de su propio historial. La significancia estadística y el valor del coeficiente AR(1) revelan una inercia considerable, congruente con una tendencia de crecimiento o declive sostenido. Además, el modelo ha identificado y cuantificado una clara estructura estacional de 12 meses mediante los coeficientes SAR(1) y SAR(2), confirmando que el nivel de empleo presenta patrones que se repiten en ciclos anuales. Finalmente, el proceso de integración del modelo señala que la serie posee una tendencia no estacionaria que fue exitosamente capturada y removida, demostrando que el modelo está bien especificado.

3.2 ARIMA BAYESIANO BSTS

El modelo Bayesian Structural Time Series (BSTS) es una técnica de pronóstico que descompone la serie de tiempo en componentes estructurales fundamentales, a diferencia de los métodos tradicionales que se centran únicamente en la correlación. El BSTS modela explícitamente la tendencia (cómo evoluciona el nivel de la serie), la estacionalidad (patrones que se repiten en el tiempo) y el impacto de covariables externas de manera dinámica. Utilizando un enfoque bayesiano, el modelo no solo proporciona una predicción puntual, sino también un intervalo de credibilidad que cuantifica la incertidumbre. Esto lo convierte en una herramienta flexible y poderosa, especialmente valiosa para inferencia causal y pronósticos robustos en entornos complejos (Scott & Varian, 2014).

La implementación de BSTS busca, mediante el enfoque bayesiano, obtener un modelo con pronósticos más adecuados, modelando la tendencia y la estacionalidad y capturando mejor los cambios bruscos en la serie mediante priors adecuadas.

En este caso el BSTS, recibe como parametros la estructura del modelo, y en cada componente recibe una prior, para este caso debido a la complejidad de la serie de tiempo se definiran los priors en default, es decir se dejan libres para que sean ajustados por el modelo, ademas, se incluirá el componente de tendencia local AddLocalLinearTrend, significa que el nivel y la pendiente se ajustan en cada paso de tiempo, permitiendo que la tendencia se acelere, desacelere o se frene, adaptándose a crisis o booms económicos; se incluirá tambien el componente estacional, debido a que tanto ARIMA, como el analisis inicial de la serie original muestran un fuerte factor periodico.

```

y <- train

ss <- list()
ss <- AddLocalLinearTrend(ss, y)
ss <- AddSeasonal(ss, y, nseasons = 12)

set.seed(123)
fit_bayesian <- bstes(y, state.specification = ss, niter = 10000, ping = 0)
summary(fit_bayesian)

## $residual.sd
## [1] 14022.14
##
## $prediction.sd
## [1] 30010.73
##
## $rsquare
## [1] 0.9995449
##
## $relative.gof
## [1] 0.1660029

fc_bayesian <- predict(fit_bayesian, horizon = h, burn = 1000)
mse_bayesian <- mean((as.numeric(test) - as.numeric(fc_bayesian$mean))^2, na.rm = TRUE)

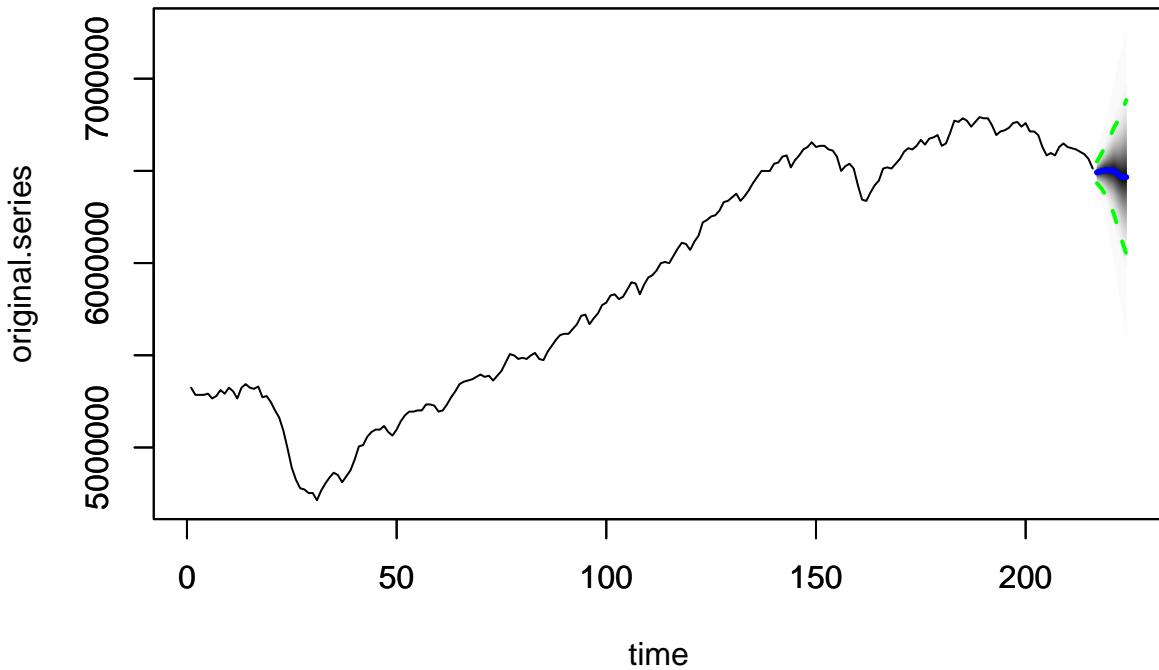
print(mse_bayesian)

## [1] 445075599

plot(fc_bayesian, main="Pronostico empleabilidad sector manufacturero Mexico 2025 BSTS")

```

Pronostico empleabilidad sector manufacturero Mexico 2025 BSTS



Resultados

Dado que la magnitud de la variable de interes esta dada en millones, los MSE de los modelos son, razonablemente, grandes por lo que se usara una funcion para determinar de manera mas sencilla el mse mas bajo.

```
comparar_mse <- function(mse_hw, mse_arima, mse_b) {  
  valores <- c(HoltWinters = mse_hw,  
                ARIMA      = mse_arima,  
                Bayesiano  = mse_b)  
  
  menor <- which.min(valores)  
  list(  
    modelo_ganador = names(valores)[menor],  
    mse_minimo     = valores[menor],  
    todos_los_mse  = valores  
)  
}  
  
resultado <- comparar_mse(mse_hw, mse_arima, mse_bayesian)  
resultado
```

```

## $modelo_ganador
## [1] "Bayesiano"
##
## $mse_minimo
## Bayesiano
## 445075599
##
## $todos_los_mse
## HoltWinters      ARIMA     Bayesiano
## 6550134140    1472929647    445075599

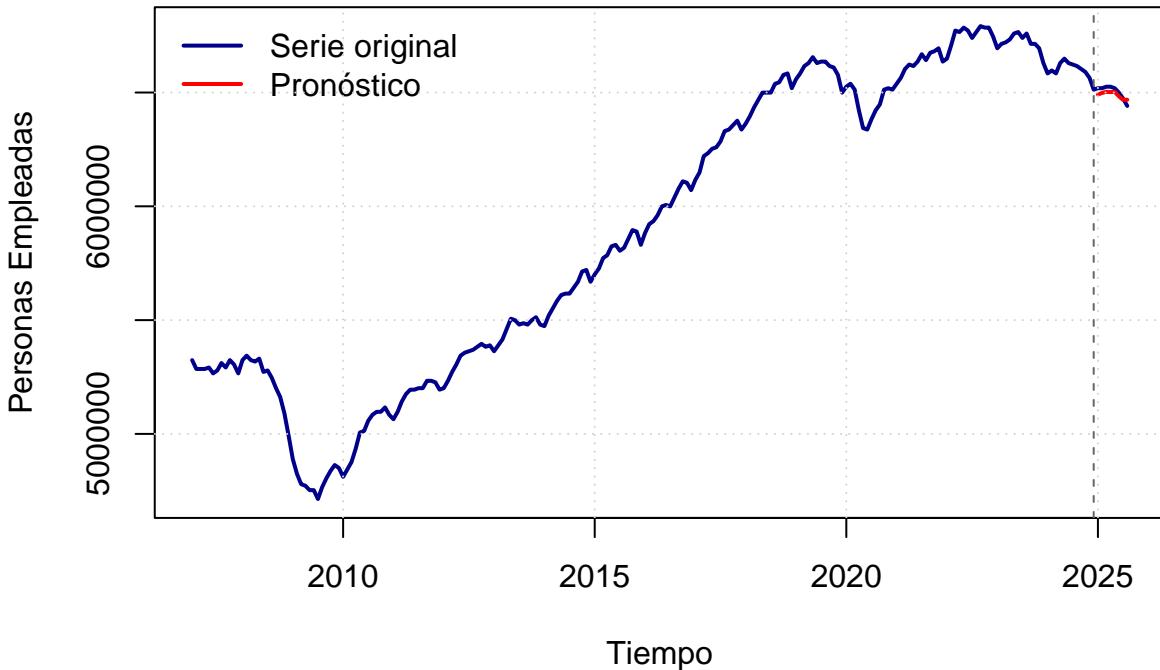
cat("Diferencia en el MSE entre el modelo ARIMA y BSTS:", mse_arima - mse_bayesian, "\n"

## Diferencia en el MSE entre el modelo ARIMA y BSTS: 1027854048

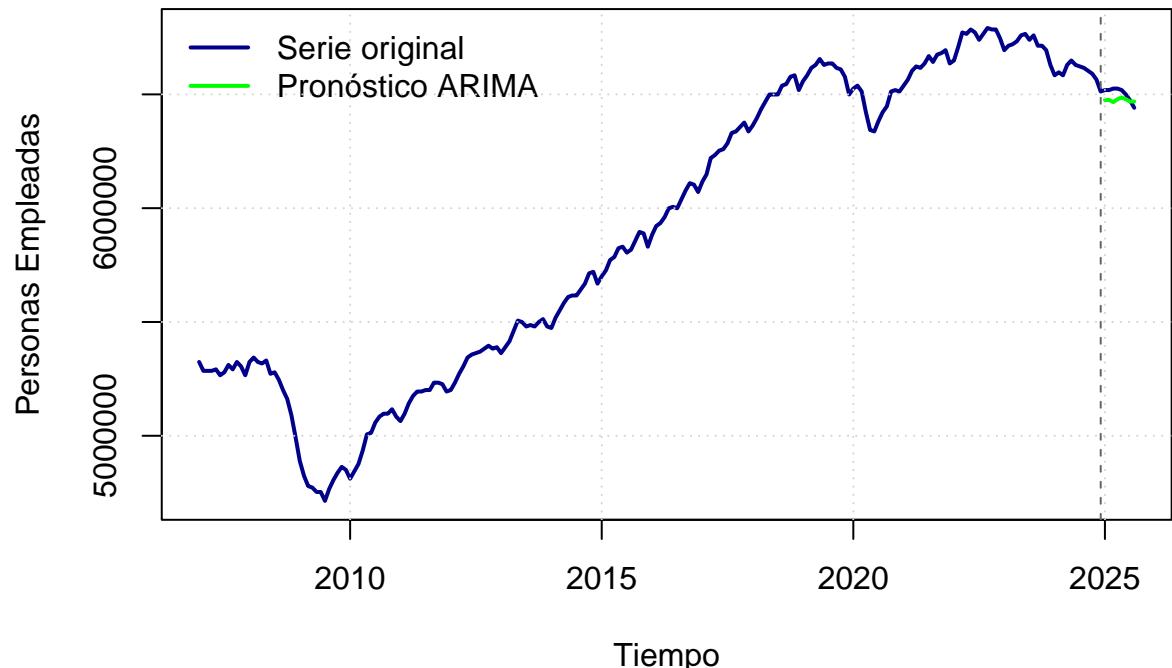
```

Comparemos los pronósticos de los dos mejores modelos, respecto a la serie de tiempo original:

Serie Original de Empleo vs. Pronóstico BSTS



Serie Original de Empleo vs. Pronóstico ARIMA



sualmente, se percibe un mejor rendimiento del modelo BSTS. Si bien ambas implementaciones ofrecen pronósticos robustos, la aproximación bayesiana demostró un rendimiento considerablemente superior.

3.4 Pronostico 2026

Como se observó en el apartado anterior, el modelo con mejor rendimiento fue indudablemente el modelo BSTS, con un desempeño superior en un orden de magnitud respecto a los demás modelos. Por lo tanto, se utilizó esta implementación para la serie de tiempo original y se pronosticó la empleabilidad en la industria manufacturera para el año 2026.

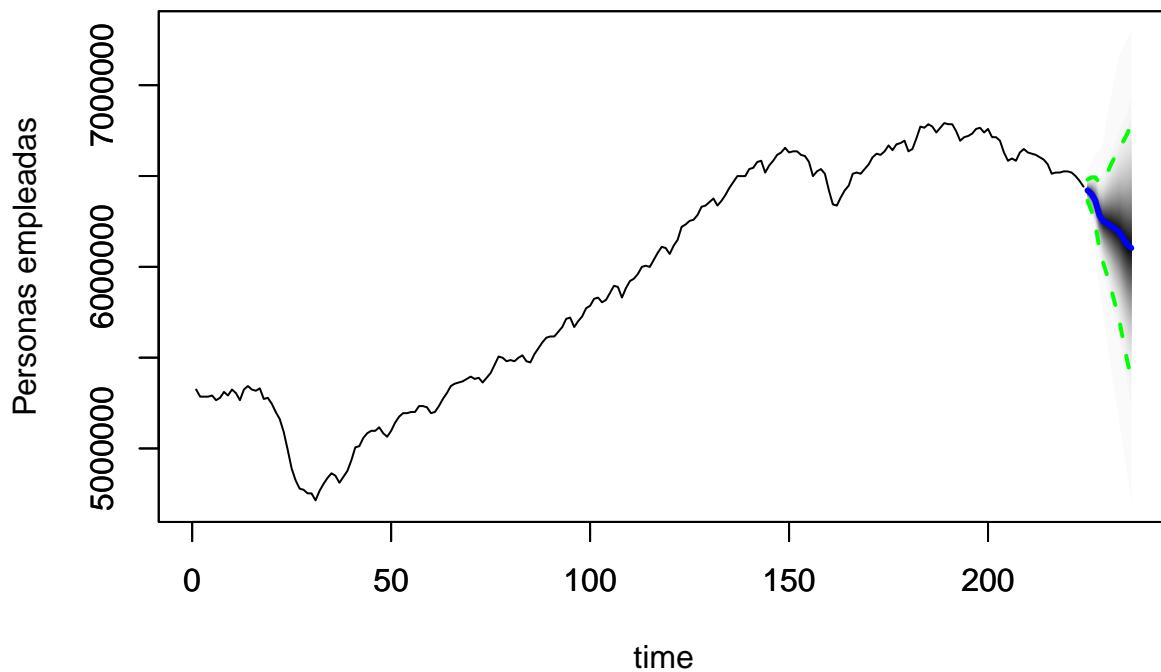
```
ss <- list()
ss <- AddLocalLinearTrend(ss, manufatura_ts)
ss <- AddSeasonal(ss, manufatura_ts, nseasons = 12)

set.seed(123)
fit <- bststs(manufatura_ts, state.specification = ss, niter = 10000, ping = 0)

fc_2026 <- predict(fit, horizon = 12, burn = 1000)
```

```
plot(fc_2026, main="Pronostico de empleabilidad Industria Manufacturera 2026 BSTS", ylab=
```

Pronostico de empleabilidad Industria Manufacturera 2026 BSTS



4. Conclusiones

El análisis del empleo en México representa un desafío complejo y multifactorial, especialmente por las dificultades inherentes a la recolección y estructuración de datos a gran escala. Desde una perspectiva general, acceder a fuentes confiables y actualizadas puede ser limitado; sin embargo, el uso de técnicas de web scraping y consumo de APIs demostró ser altamente efectivo para la obtención de información temporal relevante. Este tipo de datos aporta una visión dinámica de la realidad económica y se consolida como un recurso de gran valor competitivo para las organizaciones que buscan fundamentar sus decisiones en evidencia empírica.

De acuerdo con la evaluación de los resultados, el modelo Bayesian Structural Time Series (BSTS) mostró el mejor desempeño predictivo dentro del periodo de prueba, superando al modelo ARIMA. Su fortaleza radica en la flexibilidad del enfoque bayesiano y en la capacidad de su estructura de espacio de estados para modelar de manera dinámica los componentes de tendencia y estacionalidad. A diferencia de los modelos de coeficientes constantes, BSTS permitió adaptarse a los cambios no lineales y la alta volatilidad del empleo manufacturero en México durante el periodo 2007–2025, ofreciendo pronósticos más estables y precisos.

Los resultados obtenidos constituyen una ventaja estratégica para las empresas de reclu-

tamiento, al permitirles anticipar picos de demanda laboral, optimizar sus procesos de captación y acercarse proactivamente a clientes del sector manufacturero antes de las temporadas de alta contratación. Esta capacidad predictiva mejora la eficiencia operativa, potencia las estrategias de CRM y eleva la probabilidad de éxito comercial, posicionando a la firma como un socio clave para la gestión de talento especializado. Además, el proceso desarrollado es replicable y escalable, pudiendo aplicarse a otros sectores o geografías mediante la incorporación de indicadores macroeconómicos y laborales. Esto abre una oportunidad significativa para desarrollar soluciones analíticas de alto impacto económico y de gran potencial de crecimiento empresarial.

5 Referencias

- Agencia de Reclutamiento. (2025). ¿Cuáles son los costos de agencia de reclutamiento en 2025? Recuperado de <https://agenciadereclutamiento.com.mx/cuales-son-los-costos-de-agencia-de-reclutamiento-en-2025/>
- Banco Mundial. (2024). Manufacturing, value added (% of GDP) – Mexico. Recuperado de <https://data.worldbank.org/indicator/NV.IND.MANF.ZS?locations=MX>
- Del Moral, S. (2025). ¿Cuánto cobra un headhunter o agencia de reclutamiento y selección? Recuperado de <https://sergiodelmoral.com/cuanto-cobra-headhunter-agencia-reclutamiento-seleccion/>
- Instituto Nacional de Estadística y Geografía (INEGI). (2024). Producto Interno Bruto y estadísticas del sector manufacturero 2023. Recuperado de <https://www.inegi.org.mx>
- Scott, S. L., & Varian, H. R. (2014). Predicting the present with Bayesian structural time series. International Journal of Mathematical Modelling and Numerical Optimisation, 5(1–2), 4–23. <https://doi.org/10.1504/IJMMNO.2014.059942>
- Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6(3), 324–342. <https://doi.org/10.1287/mnsc.6.3.324>
- Box, G. E. P., & Jenkins, G. M. (1970). Time Series Analysis: Forecasting and Control. San Francisco, CA: Holden-Day.