

Maestría en Ciencia de Datos Universidad Icesi

Fundamentos de Analítica de Datos II - Taller No. 1

Daniel Martínez Villegas

Luis Felipe Montenegro Fuentes

Teniendo en cuenta los datos de empleo provistos por el DANE, en este documento se llevará a cabo un ejercicio enfocado en utilizar métodos estadísticos aplicados a series temporales para pronosticar el número de ocupados en las 13 principales ciudades de Colombia. Según los datos disponibles, se tiene información de la variable para un rango que va de enero de 2001 a abril de 2019. Dicho eso, el objetivo principal es extrapolar la información para los seis meses siguientes (may-jun).

Protocolo de Evaluación & Elección del Modelo

Gráfica 1. Protocolo de evaluación utilizando el modelo Holt Winters

Fuente: DANE – Elaboración propia

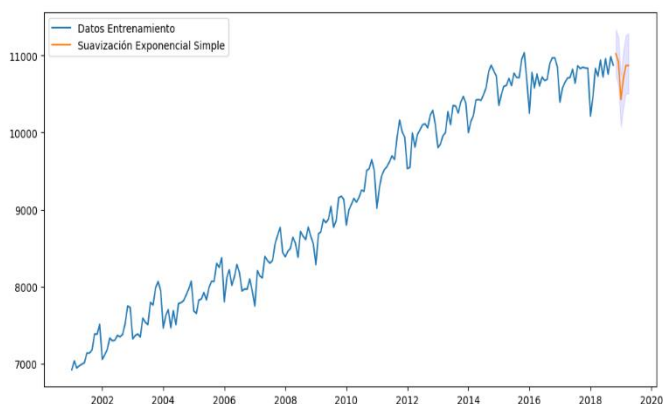


Tabla 1. Elección del mejor modelo según el RMSE

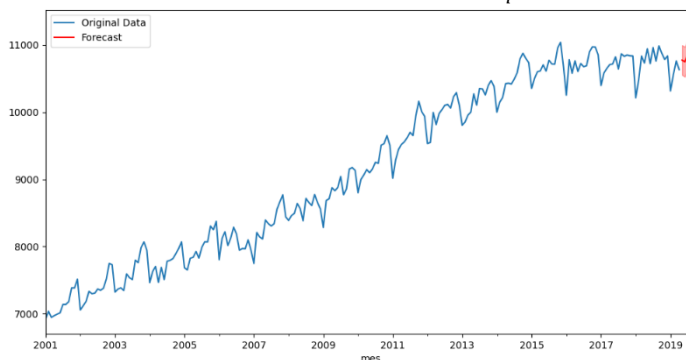
Fuente: DANE – Elaboración propia

| | RMSE |
|--|------------|
| Modelo | |
| suavizacion_exp_lin_holt_winters_add_add_mul_mse | 156.414707 |
| suavizacion_exp_lin_holt_winters_mul_mul_mul_mse | 160.567634 |
| suavizacion_exp_lin_holt_winters_add_add_add_mse | 163.608148 |
| suavizacion_exp_lin_holt_winters_mul_add_add_mse | 166.390499 |
| suavizacion_exp_lin_holt_winters_add_mul_add_mse | 169.239858 |
| suavizacion_exp_lin_holt_winters_mul_mul_add_mse | 208.710500 |
| suavizacion_exp_simple_mul_mse | 300.428295 |
| suavizacion_exp_simple_add_mse | 300.429030 |
| suavizacion_exp_lin_holt_add_add_mse | 370.157975 |
| suavizacion_exp_lin_holt_mul_add_mse | 371.544361 |
| suavizacion_exp_lin_holt_mul_mul_mse | 381.545888 |
| suavizacion_exp_lin_holt_add_mul_mse | 381.545888 |

Pronostico

Gráfica 2. Pronostico del número de ocupados en las 13 principales ciudades (may-jun 2019)

Fuente: DANE – Elaboración Propia



Teniendo en cuenta los datos disponibles, el modelo plantea que para los meses de mayo a junio de 2019 se presentaría un aumento en el número de ocupados, sin embargo, es importante mencionar que la exactitud de dicha extrapolación podría verse afectada o limitada por choques externos en el entorno económico (crisis económicas, políticas gubernamentales, etc) que impacten de manera importante la senda de crecimiento de la variable analizada.