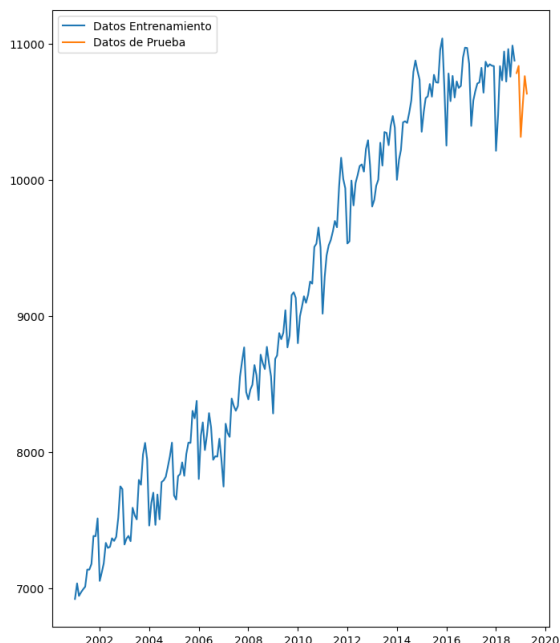


PRONOSTICO – PERSONAS OCUPADAS DE 13 CIUDADES PRINCIPALES DEL PAÍS / Raul Echeverry – Esteban Ordoñez

Los datos corresponden al número de personas ocupadas de 13 ciudades principales del país, que comprenden desde enero de 2001 hasta abril de 2019. El objetivo es pronosticar la cantidad de personas ocupadas en los meses de mayo a octubre de 2019, a través del uso de modelos de regresión lineal y uso de funciones polinómicas de distintos grados, considerando estacionalidad y sin estacionalidad.

Protocolo de evaluación

Para lograr pronosticar los 6 meses siguientes a abril 2019, se realiza un protocolo de evaluación para entrenar y escoger el mejor modelo, esto implica usar los últimos seis meses, es decir desde nov 2018 a abril 2019.

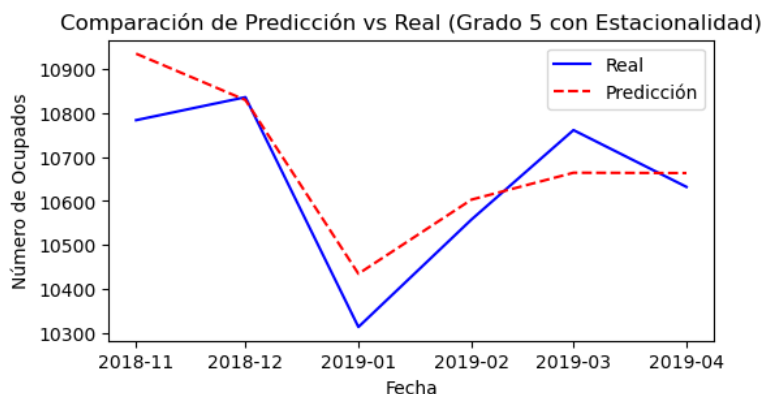


Selección de Modelo

Se prueban distintos modelos, iniciando con regresión lineal hasta polinomio de grado 10, tanto con estacionalidad como sin estacionalidad, los RMSE son los siguientes:

Grado de Polinomio	Sin Estacionalidad	Con Estacionalidad
1 (Regresión Lineal)	718,65	657,51
2	646,60	590,48
3	238,47	130,74
4	222,84	221,62
5	219,95	91,01
6	337,13	225,18
7	3.709,35	3.709,35
8	6.086,66	6.086,66
9	8.655,08	8.655,08
10	13.431,66	13.431,66

Datos reales Vs Predicciones del mejor modelo



Pronóstico de la cantidad de personas ocupadas en los meses de junio a noviembre de 2019

Fecha	Pronostico
01/06/2019	10.719,88
01/07/2019	10.676,13
01/08/2019	10.749,27
01/09/2019	10.721,44
01/10/2019	10.778,77

Comparado con los modelos de medias móviles, suavización exponencial simple, lineal de Holt, y lineal de Winters (Holt-Winters), cuyo mejor RMSE fue 93.5, este nuevo enfoque logró una ligera mejora con un RMSE de 91.01.

Limitaciones:

El modelo de regresión polinómica de grado 5 con estacionalidad presenta limitaciones significativas, ya que, aunque los residuos cumplen con el supuesto de normalidad, no cumplen con los supuestos de ruido blanco, mostrando autocorrelación y heterocedasticidad. Esto sugiere que el modelo no captura adecuadamente todos los patrones en los datos y que la varianza de los errores no es constante, lo que puede comprometer la precisión de las predicciones y la validez de las inferencias.