

Reconocimiento de Patrones Temporales en Series de Tiempo para la Predicción



Anderson Brian Flores Suaña

^{1,2,3}Facultad de Ingeniería Estadística e Informática, Universidad Nacional del Altiplano

1. RESUMEN

Este estudio compara tres modelos de series temporales (Prophet, ARIMA y LSTM) para predecir brotes de dengue en Perú usando datos 2020-2023. Se identificaron patrones temporales y tendencias epidemiológicas mediante análisis estadístico y modelado predictivo. ARIMA presentó el mejor rendimiento general con $R^2 = 0.96$, seguido por LSTM ($R^2 = 0.72$) y Prophet ($R^2 = 0.44$). Los modelos permiten anticipar períodos de alto riesgo epidemiológico y optimizar la vigilancia en salud pública.

Palabras clave: ARIMA, dengue, LSTM, predicción epidemiológica, Prophet, series temporales.

2. METODOLOGÍA

Datos: Casos de dengue reportados en Perú (2020-2023) obtenidos del MINSA.

Preprocesamiento:

- Agrupación semanal de datos diarios
- Tratamiento de valores atípicos (método Tukey)
- Interpolación lineal para valores faltantes
- Transformación logarítmica para estabilizar varianza

Modelos Implementados:

- Prophet:** Descomposición bayesiana con componentes de tendencia y estacionalidad
- ARIMA(2,1,2):** Modelo autorregresivo integrado de media móvil
- LSTM:** Red neuronal recurrente con 50 unidades y ventana temporal de 4 semanas

Evaluación: Validación temporal (80/20), métricas RMSE, MAE, MAPE y R^2 .

3. ANÁLISIS DE DATOS

Total de casos (2020-2023): 412,580

Distribución por Edad:

Grupo Etario	Casos	%
0-11 meses	2,393	0.6%
1-4 años	15,208	3.7%
5-14 años	80,597	19.5%
15-29 años	121,805	29.5%
30-59 años	155,216	37.6%
60+ años	37,361	9.1%

Departamentos más afectados:

- Piura: 95,651 casos
- Lambayeque: 35,257 casos
- Lima: 33,213 casos
- Ica: 33,325 casos
- Loreto: 28,423 casos

Estacionalidad: Pico máximo en mayo 2023 con 53,510 casos. Patrón estacional marcado con mayor incidencia entre marzo-junio.

RESULTADOS COMPARATIVO

Métricas de Rendimiento:

Modelo	RMSE	MAE	MAPE	R^2
ARIMA	556.60	221.29	11.15%	0.96
LSTM	2,663.19	1,553.03	19.20%	0.72
Prophet	2,220.03	1,462.15	116.98%	0.44

Predicciones 2024 (Enero-Marzo):

Mes	ARIMA	LSTM	Prophet
Enero	1,022	1,083	1,380
Febrero	936	994	1,350
Marzo	908	878	1,320

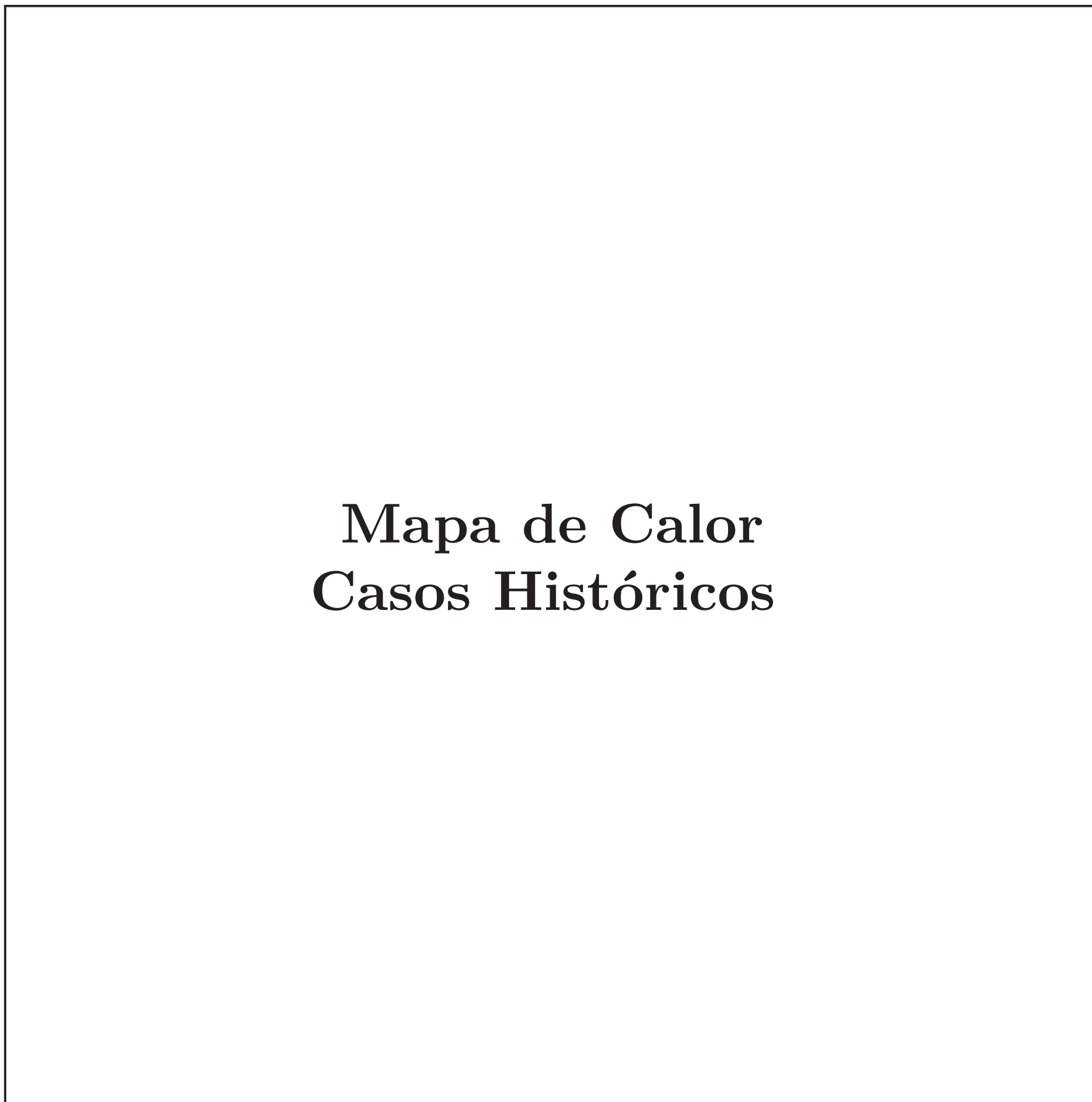
Conclusión de Modelos:

- ARIMA:** Mejor precisión general
- LSTM:** Captura patrones complejos
- Prophet:** Detecta estacionalidad pero con menor precisión

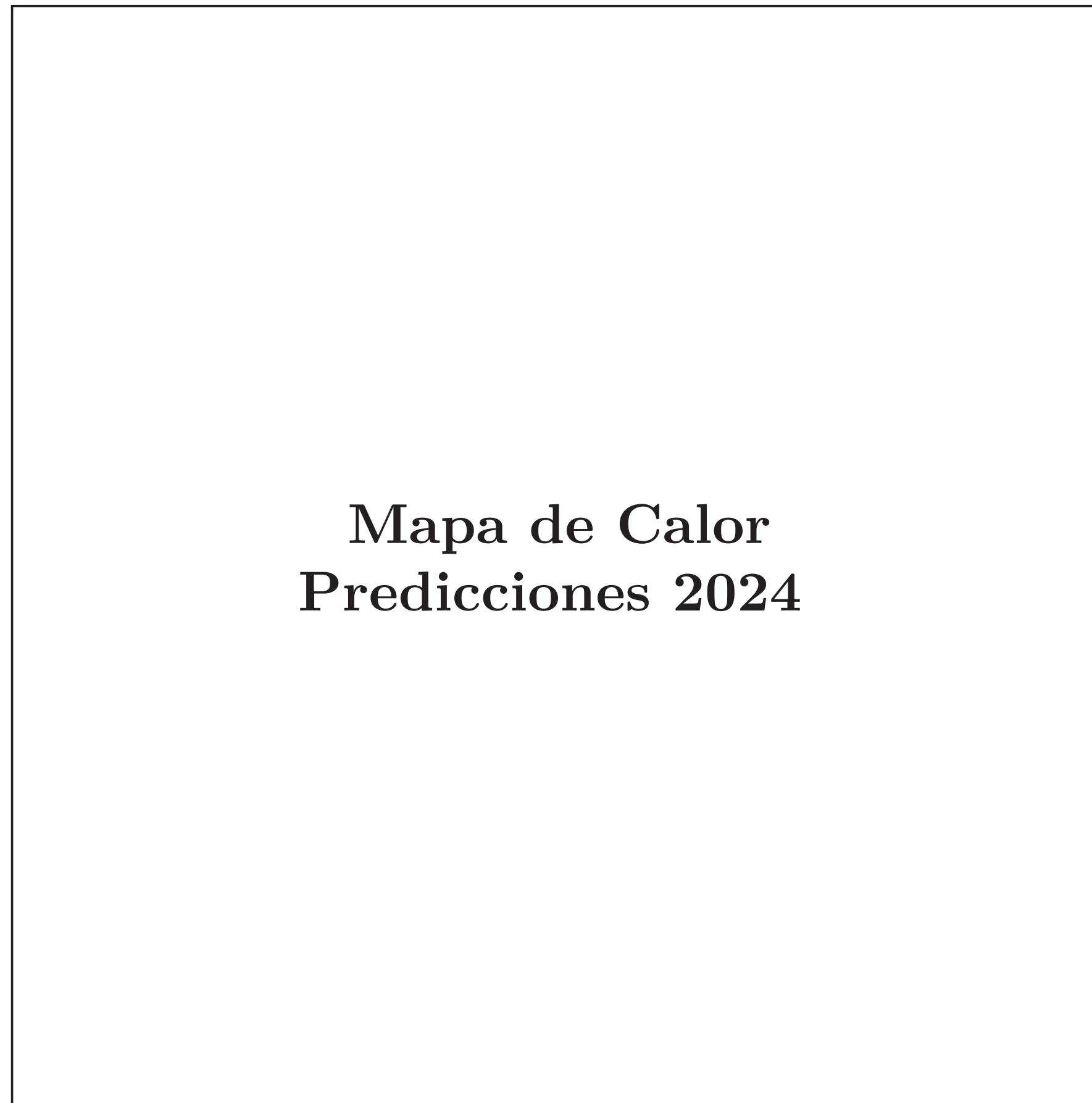
Gráfico de Predicciones Temporales

5. DISTRIBUCIÓN GEOGRÁFICA

Mapa 1: Casos Totales por Departamento (2020-2023)



Mapa 2: Predicción por Departamento (2024)



Análisis Regional - Predicciones ARIMA 2024 (12 semanas):

Departamento	Histórico	Predicción	Departamento	Histórico	Predicción
San Martín	20,647	1,954	Ucayali	25,937	845
Piura	95,651	1,945	Ica	33,325	671
La Libertad	27,316	1,846	Junín	16,565	612
Lambayeque	35,257	912	Loreto	28,423	816
Ancash	13,831	933	Tumbes	18,161	271

Observaciones Regionales: Costa norte (Piura, Lambayeque) y selva (San Martín) mantienen mayor riesgo epidemiológico. Sierra presenta baja incidencia general.

6. CONCLUSIONES

- Modelo Óptimo:** ARIMA(2,1,2) demostró superior rendimiento con $R^2 = 0.96$ y MAPE = 11.15% para predicción epidemiológica.
- Estacionalidad:** Se identificó patrón estacional marcado con picos mayo-junio y valle agosto-octubre.
- Distribución Regional:** Costa norte y selva mantienen mayor riesgo epidemiológico.
- Herramienta Efectiva:** Los modelos de series temporales son útiles para vigilancia epidemiológica y planificación sanitaria.
- Impacto:** Permite implementar medidas preventivas oportunas y optimizar recursos de salud pública.

Recomendaciones:

- Implementar monitoreo continuo
- Incorporar variables climáticas
- Desarrollar sistema de alerta temprana
- Validar con datos externos

7. REFERENCIAS

- Dostal, T., Meisner, J., et al. (2022). "The effect of weather and climate on dengue outbreak risk in Peru, 2000–2018: A time-series analysis". *PLOS Neglected Tropical Diseases*.
- Goche, K.S.R., Castro, M.V.L., et al. (2025). "Epidemiological dynamics of dengue in Peru: Temporal and spatial drivers between 2000 and 2022". *PLOS ONE*.
- Attanayake, A.M.C.H. & Perera, S.S.N. (2021). "Time series analysis for modeling the transmission of