

# Estrategias de Entrenamiento de Modelos de Predicción en Redes Celulares con Segmentación

J.A. Villegas, C. Gijón, M. Toril, S. Luna-Ramírez, M. Fernández

{jvillegasmartin, cgm, mtoril, sluna, mariano}@ic.uma.es

Telecommunication Research Institute (TELMA), Universidad de Málaga, Málaga, España.

**Resumen**—In 5G and beyond systems, Network Slicing (NS) enables the deployment of multiple logical networks customised for specific verticals over a common physical infrastructure. In the radio access network, mobile operators need models to forecast slice performance for an efficient and proactive slice redimensioning. This work tackles slice throughput forecasting in the short term through supervised learning, namely Time Series Mixer-Only algorithm. Four training strategies are assessed, relying on slice throughput time series, synthetic time series, or a combination of both. Performance assessment is carried out with a realistic dataset generated via simulation in a network with 4 slices offering services with different traffic model and performance requirements to both terrestrial and aerial users. Results show that the use of transfer learning to combine synthetic and throughput data from the corresponding segment is a promising approach, with a better performance than *Random Walk* method for a time horizon of 30 days in all slices.

## I. INTRODUCCIÓN

La tecnología 5G tiene como objetivo expandir las redes de comunicaciones móviles para ofrecer una amplia gama de servicios como la transmisión de vídeo en calidad 4K, la conducción autónoma o las aplicaciones agrícolas. Estos nuevos servicios plantean requisitos de rendimiento muy heterogéneos, como la necesidad de un alto ancho de banda (*Enhanced Mobile Broadband*, eMBB), el soporte para un gran número de dispositivos limitados en coste y energía (*Massive Machine Type Communications*, mMTC) o conexiones ultra-fiables con baja latencia (*Ultra-Reliable Low Latency Communications*, uRLLC). La función de segmentación de red (*Network Slicing*, NS) permite la coexistencia de los servicios al dividir la infraestructura física en redes lógicas adaptadas a servicios específicos o industrias verticales [1].

En las redes con NS, surgen nuevas funciones de red como los repartidores de capacidad que deben optimizar la asignación de recursos por segmento extremo a extremo para proporcionar una adecuada calidad de servicio [2]. Esto, junto a la activación y desactivación de segmentos, hace que la gestión de las redes con NS sea extremadamente compleja [1]. Por tanto, para asegurar el cumplimiento de los acuerdos de servicio, las funciones de red deben operar de manera proactiva, anticipándose a posibles degradaciones [3]. Para ello, es crucial predecir indicadores de rendimiento (*Key Performance Indicators*, KPIs) de los segmentos a corto plazo, es decir, en un horizonte temporal de segundos. Con los últimos avances en inteligencia artificial, esta tarea se puede abordar a través del uso de técnicas de aprendizaje supervisado (*Supervised Learning*, SL).

En la literatura, la predicción del rendimiento de celda se ha abordado habitualmente mediante el uso de modelos basados en redes neuronales con arquitecturas convolucionales, con

memoria (*Long Short-Term Memory*, LSTM) o transformadores. Estas técnicas han sido extendidos para predecir el rendimiento de segmento (p.ej., [4], [5], [6]), pero, ante la inexistencia de juegos de datos públicos con medidas históricas del rendimiento de redes 5G con NS, estas contribuciones utilizan juegos de datos recopilados en redes heredadas.

La técnica de predicción más adecuada para un caso de uso concreto (resolución temporal de los datos, KPI a predecir...) depende de las propiedades de las series temporales implicadas (ruido, estacionalidad, tendencia, presencia de escalones...). En [7], se demostró que algunas de estas propiedades varían significativamente en los distintos segmentos que operan en una red, sugiriendo que es necesario entrenar un modelo de predicción específico por segmento. Esta premisa aún no ha sido comprobada. Para realizar este estudio, se requiere un juego de datos recopilados por segmento y no por celda, como en contribuciones anteriores.

Este trabajo aborda la predicción de KPIs a nivel de segmento mediante el uso de SL empleando por primera vez un juego de datos que contiene medidas del rendimiento por segmento. Dicho juego de datos se ha obtenido con un simulador que emula la actividad de una red de acceso radio segmentada en un escenario realista. Las principales contribuciones son: a) la evaluación del desempeño de un modelo de SL para predecir el *throughput* de varios segmentos a corto plazo, y b) la comparativa de diversas estrategias de entrenamiento de dicho modelo de predicción, basadas en el uso de datos reales, sintéticos, y en combinaciones de ambos a través del uso de aprendizaje por transferencia.

Este documento se estructura como sigue. La sección II presenta la metodología de predicción empleada. La sección III expone las pruebas realizadas. Por último, la sección IV resume las principales conclusiones.

## II. METODOLOGÍA DE PREDICCIÓN

Asúmase una red móvil 5G donde varios servicios operan simultáneamente. El objetivo de este trabajo es encontrar la mejor estrategia de entrenamiento para predecir el rendimiento de cada segmento a corto plazo, es decir, con un horizonte temporal de segundos empleando SL. Se emplea esta técnica de aprendizaje porque se cuenta con un juego de datos etiquetado en el que se conoce exactamente la salida que tendrán las entradas. Esta sección primero presenta el modelo de SL considerado y después define las estrategias de entrenamiento comparadas.

### A. Modelo de predicción

El análisis se realiza con el modelo de predicción *TsMixer-Only* (Time Series Mixer - Only), vencedor de la competición

de predicción M5 en el año 2020 [8].

TsMix-Only es un modelo de SL desarrollado específicamente para la predicción de series temporales. El modelo es capaz de capturar patrones temporales persistentes, como tendencias o estacionalidades. Su arquitectura se basa en el apilamiento de modelos que asumen relaciones lineales y perceptrones multicapa (*MultiLayer Perceptron*, MLP), que le permite capturar relaciones temporales. Además, el uso de técnicas como la normalización o conexiones residuales facilitan el aprendizaje, mejorando la eficiencia computacional y reduciendo el número de parámetros entrenables [9].

### B. Estrategias de entrenamiento

En este trabajo, el modelo TsMix-Only se entrena empleando 4 estrategias descritas a continuación:

- E1: modelo global entrenado con datos reales. Esta estrategia, utilizada como caso base, consiste en entrenar un único modelo de predicción haciendo uso de series temporales de un KPI procedente de todos los segmentos de la red, siendo explotado de forma independiente en cada segmento. Las potenciales ventajas son la necesidad de entrenar solo un modelo y la capacidad de generalización al ser dicho modelo entrenado con series temporales procedentes de segmentos con distinta naturaleza.
- E2: modelos específicos entrenados con datos reales. El estudio realizado en [7] mostró que las series temporales de los segmentos que operan simultáneamente sobre una misma red presentan propiedades distintas. Esta estrategia consiste en entrenar un modelo específico para cada segmento empleando series temporales del mismo. Su potencial ventaja es la capacidad de capturar las peculiaridades de la evolución temporal del rendimiento de un segmento concreto, a cambio de una posible pérdida de generalización.
- E3: modelo global entrenado con datos sintéticos. En las redes móviles, la distribución uniforme del tráfico o la existencia de condiciones radio similares en las celdas en las que se recopilan datos históricos pueden ocasionar que todas las series temporales disponibles para entrenar un modelo de predicción sean similares. Esto se traduce en una pérdida de generalidad del modelo, que funcionará de manera subóptima ante series temporales con otro tipo de propiedades (p.ej., tendencia o estacionalidad). Para solventar este problema, esta estrategia entrena un modelo de predicción global empleando series sintéticas que contienen un gran espectro de tendencias, ruido, estacionalidades y cambios abruptos de nivel, que permitirán aumentar la generalidad. Como desventaja, este modelo, que podría emplearse en cualquier caso de uso de predicción, no captará las peculiaridades de la evolución del rendimiento de ningún segmento.
- E4: modelos específicos entrenados con aprendizaje por transferencia. Esta última estrategia consiste en entrenar un modelo de predicción específico por segmento que combine las ventajas de las estrategias E2 y E3 mediante el uso de aprendizaje por transferencia. Así, en primer lugar, se entrena un modelo genérico descrito en E3, aprendiendo a predecir patrones generales. Tras ello, los pesos de este modelo se reentrenan reduciendo el ratio de aprendizaje para cada segmento tal y como se

TABLA I: Características de los servicios.

| Segmento        | Modelo de tráfico                                    | Usuario              |
|-----------------|--|----------------------|
| S1<br>(FTP)     | Full buffer<br>Duración: 15 min                      | Aéreo<br>5 km/h      |
| S2<br>(VIDEO)   | Traza H.264/MPEG-4 AVC<br>Duración: U [30, 300] s    | Terrestre<br>0 km/h  |
| S3<br>(HAPTIC)  | Full buffer<br>Duración: U [300, 600] s              | Terrestre<br>0 km/h  |
| S4<br>(DRIVING) | 1 paq. 201 B cada 100 ms<br>Duración: U [300, 600] s | Terrestre<br>50 km/h |

describe en E2, de manera que el modelo resultante para cada segmento pueda captar peculiaridades del mismo no presentes en el juego de series sintéticas.

## III. ANÁLISIS DEL RENDIMIENTO

Esta sección detalla los experimentos realizados. Por claridad, primero se presentan los juegos de datos, luego se describe la metodología experimental y por último los resultados.

### A. Juegos de datos

A continuación, se presentan los juegos de datos empleados.

1. *Series temporales de throughput de segmento*: el primer juego de datos empleado contiene datos históricos del throughput agregado por segmento en cada celda de una red. Dicho juego de datos se ha generado empleando la herramienta de simulación descrita en [10]. Se trata de un simulador dinámico de nivel de sistema que emula la actividad del enlace descendente de una red de acceso radio realista 5G que cuenta con la funcionalidad de NS. Para obtener una descripción completa de la herramienta (no incluida aquí por brevedad), se refiere al lector a [10]. En el simulador, se configura un escenario compuesto por 108 macroceldas con ubicaciones distribuidas en un entorno suburbano. Las celdas funcionan a una frecuencia de 2,1 GHz con un ancho de banda de 10 MHz. En la red, 4 segmentos operan simultáneamente para servir a usuarios caracterizados por distintos terminales (*smartphones* o drones), aplicaciones y patrones de movilidad. La tabla I describe cada segmento. S1 cursa el tráfico de una flota de drones en movilidad que demandan un servicio de descarga de ficheros (*File Transfer Protocol*, FTP) muy pesados (es decir, sus conexiones son muy largas). S2 y S3 sirven a usuarios estáticos terrestres. Los usuarios de S2 demandan un servicio de reproducción de vídeo en directo (VIDEO), también de tipo eMBB, pero con conexiones más cortas. En cambio, S3 cursa el tráfico de comunicaciones hápticas (HAPTIC), de tipo uRLLC. Por último, S4 sirve a los usuarios de conducción autónoma (DRIVING) que viajan en vehículos a alta velocidad. Este servicio es de tipo uRLLC, pero demanda un menor throughput que HAPTIC.

Para crear el juego de datos, se realizan 6 simulaciones de una duración de 19 min con una resolución temporal de 10 ms. Cada simulación presenta la misma distribución espacial relativa de los usuarios y la mezcla de servicios por celda, mientras que difieren en el nivel de carga en la red. Durante las simulaciones, se almacena el throughput agregado por celda y segmento con una resolución temporal de 1 s, calculado como el número de bits transmitidos en la capa PDCP (*Packet Data Convergence Protocol*) por unidad de tiempo para los recursos físicos asignados a un segmento en

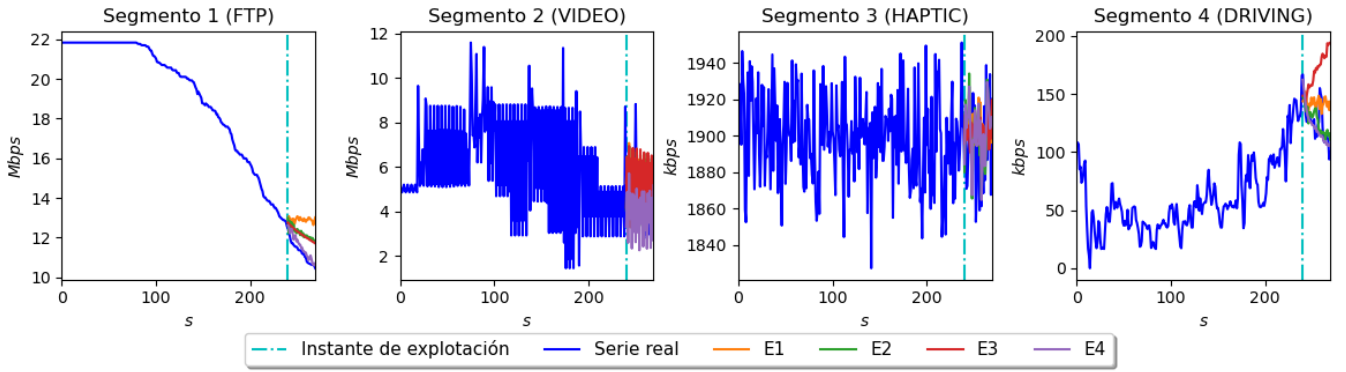


Fig. 1: Predicciones sobre una celda

TABLA II: Hiperparámetros del modelo.

| Hiperparámetro               | Valor           |
|------------------------------|-----------------|
| Tasa de <i>dropout</i>       | 0.1 - 0.3 - 0.5 |
| Función de activación capa X | Lineal          |
| Tasa de aprendizaje          | 0.01 - 0.0001   |
| Tamaño del lote              | 2048            |
| Número de bloques            | 1               |

el DL de una celda. Acto seguido, se realiza un proceso de aumento de datos reduciendo las series a una duración de 4 min y 30 s obteniendo 117.050 series temporales para el entrenamiento y 25.014 para la explotación.

2. *Series temporales sintéticas*: Para crear el juego de datos sintético, se emplea una rutina que genera 100.000 series temporales con una longitud configurable y valores comprendidos entre 0 y 1. Las series generadas presentan distintas propiedades en cuanto a: a) tendencias (nivel estable, creciente, decreciente y con cambios de tendencia), b) cambios abruptos de nivel, permanentes y transitorios, c) nivel de ruido Gaussiano, y, d) estacionalidad (series no estacionales, y estacionales con distintos periodos y patrones).

### B. Metodología experimental

Se entrenan modelos para predecir el throughput de cada segmento considerando una ventana de observación de 240 s y un horizonte de 30 s haciendo uso de las 4 estrategias de entrenamiento. En total, se entrenan 10 modelos de predicción (1 para E1 + 4 para E2 + 1 para E3 + 4 para E4).

La tabla II resume la configuración de hiperparámetros seleccionada. Los principales valores se optimizan para cada modelo mediante un barrido en el espacio de soluciones. Además, se emplea una función de activación lineal debido a que se ha comprobado que aumentar la complejidad produce resultados más sensibles al ruido o a valores atípicos. Para evitar que las series con valores mayores dominen el entrenamiento, se emplea un escalador estándar.

El desempeño de los modelos para cada segmento se mide empleando como cifra de mérito el error absoluto medio de escalamiento, *MASE*, que compara el error del modelo de predicción con el que cometería el método de pronóstico ingenuo o (*Random Walk*), que consiste en repetir la muestra más reciente. La *MASE* se calcula como

$$MASE[\%] = \frac{\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|}{\frac{1}{m-1} \sum_{j=2}^m |y_j - y_{j-1}|}, \quad (1)$$

donde  $n$  es el número de observaciones,  $y_i$  son los valores reales,  $\hat{y}_i$  son los valores predichos por el modelo y  $m$  es el tamaño de la ventana de observación. Valores de *MASE* superiores a 1 indican que las predicciones tienen un error mayor que el obtenido si se replica la muestra temporal más reciente en todos los valores del horizonte [11].

### C. Resultados

La figura 1 muestra la evolución del throughput de los 4 segmentos considerados en una celda aleatoria, así como las predicciones realizadas por las distintas estrategias de entrenamiento sobre esas series cuando el modelo se explota en el instante indicado por la línea vertical. Comparando los datos reales (curvas azules) de distintos segmentos, se observa que los distintos servicios ofrecidos en cada segmento condicionan notablemente las propiedades de las series temporales de throughput medido con resolución temporal de 1 s. En concreto, en S1 (FTP), el throughput muestra una clara tendencia debida a la variación de condiciones radio provocadas por el movimiento de los drones y no presenta estacionalidad, ya que el tráfico FTP es de tipo *full buffer*. En S2 (VIDEO), el throughput presenta una estacionalidad causada por el tráfico a ráfagas. No existe tendencia, ya que los usuarios son estáticos, pero sí cambios de nivel debido al comienzo o fin de conexiones. La principal propiedad de la evolución del throughput en S3 (HAPTIC) es la presencia de mucho ruido, probablemente provocado por la asignación de recursos radio a distintos usuarios en instantes temporales consecutivos. Finalmente, el throughput en S4 (DRIVING) tiene una pendiente impredecible (comportamiento tipo *Random Walk*) debido a la alta movilidad de los usuarios, que provoca continuos traspasos. Además, la serie temporal presenta cierto ruido debido a cambios bruscos en las condiciones radio de los usuarios por desvanecimientos rápidos. Estas claras diferencias entre segmentos reflejan la complejidad y heterogeneidad de segmentos presentes en las redes B5G, justificando el estudio realizado en este trabajo.

La figura 2 presenta la evolución del indicador *MASE* a lo largo del horizonte de predicción para todos los segmentos y estrategias de entrenamiento consideradas. Comparando las curvas de E1 (naranja) y E2 (verde), se observa que el modelo específico por segmento de E2 tiene un desempeño similar (en S3) o peor (en el resto de segmentos) que el modelo multi-segmento de E1. Por tanto, se puede concluir que, cuando el modelo de predicción se entrena con datos reales, conviene crear un modelo específico para cada segmento, que capte las

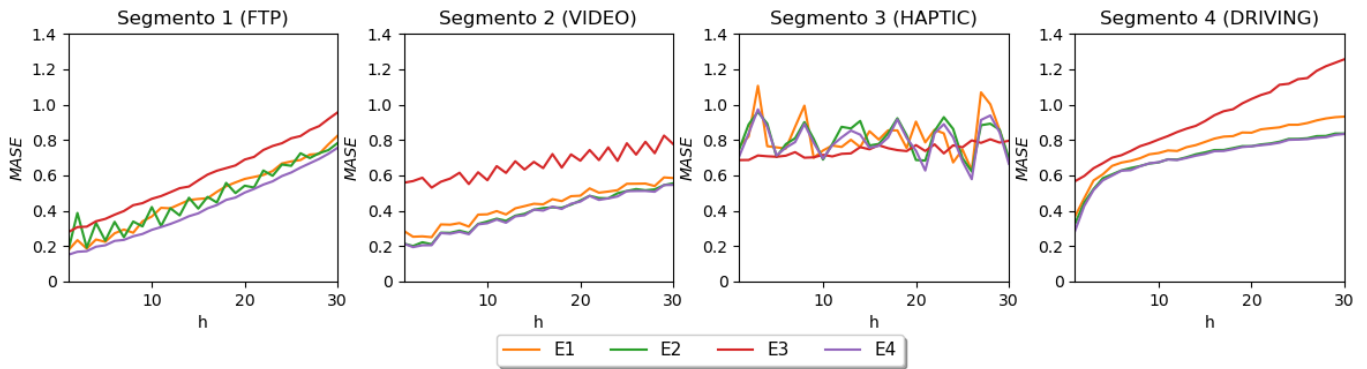


Fig. 2: Evolución de las cifras de mérito para un horizonte de 30 s.

particularidades de la evolución de sus KPIs. Por otro lado, E3 solo muestra una precisión comparable al resto de modelos en S3 (en el resto de modelos, obtiene valores de  $MASE > 1$ ). Esto indica que, aunque un modelo genérico entrenado con series sintéticas no funciona bien con series temporales con patrones complejos, como las series a predecir en este trabajo, este modelo sí que tiene capacidad para eliminar el ruido presente en las series de S3. En cambio, cuando el modelo creado en E3 se reentrena con datos de throughput por segmento en la estrategia E4, el desempeño es siempre similar o mejor con respecto a E2, especialmente en S1 ( $MASE$  de 0.47 para E2, y de 0.41 para E4 en las predicciones a 15 segundos vista). Esto revela que los modelos entrenados en E4 pueden predecir tendencias gracias al entrenamiento con un juego de datos sintético con tendencias balanceadas y captar las peculiaridades de cada segmento gracias al reentrenamiento específico por segmento con datos reales del mismo. Dichas conclusiones las avalan las predicciones de las series mostradas en la figura 1, donde las curvas lilas de E4 muestran siempre resultados razonables como por ejemplo en la serie de S1, E1 no predice bien el nivel, y E2 y E3 predicen mal la tendencia.

#### IV. CONCLUSIONES

En este trabajo, se ha abordado la predicción de KPIs agregados por celda y segmento en redes B5G con NS. En concreto, se ha evaluado el desempeño del modelo TsMix-Only para predecir el throughput de segmento en un horizonte temporal de segundos considerando varias estrategias de entrenamiento basadas en el uso de datos reales y/o sintéticos. El juego de datos de throughput en distintos segmentos se ha obtenido de un simulador que emula la actividad de una red de acceso radio segmentada en un escenario realista en el que 4 segmentos sirven tráfico generado por los servicios de FTP, vídeo, comunicaciones hápticas y conducción autónoma, respectivamente.

Un análisis del juego de datos ha mostrado que la evolución del throughput de los distintos segmentos que operan sobre una misma red a corto plazo presenta distintas propiedades de tendencia, estacionalidad o ruido, entre otras, que son muy dependientes del tipo servicio. De hecho, la mejor estrategia de entrenamiento según la  $MASE$  depende del segmento. El uso de aprendizaje por transferencia para combinar datos sintéticos y datos del correspondiente segmento (E4) ha mostrado el mejor resultado, especialmente en segmentos con KPIs cuya tendencia es predecible. En el resto de segmentos,

los modelos entrenados exclusivamente con datos del segmento (E2) muestran un comportamiento similar. En todos los segmentos, el mejor modelo ha obtenido una  $MASE$  inferior a 1, es decir, un mejor comportamiento que el método *Random Walk* en un horizonte temporal de 30 segundos.

Como línea de trabajo futura, se empleará aprendizaje multi-tarea en la fase de entrenamiento con datos reales para predecir numerosos KPIs de un segmento simultáneamente, aprovechando la correlación existente entre ellos para mejorar la robustez de las predicciones.

#### AGRADECIMIENTOS

Este trabajo lo financia la Universidad de Málaga, la Unión Europea y el Ministerio de Ciencia e Innovación (PID2021-122217OB-I00/AEI/10.13039/501100011033).

#### REFERENCIAS

- [1] Liu, Y., Clerckx, B., and Popovski, P. (2023). *Network slicing for eMBB, URLLC, and mMTC: An uplink rate-splitting multiple access approach*. *IEEE Transactions on Wireless Communications*.
- [2] Shu, Z., and Taleb, T. (2020). *A novel QoS framework for network slicing in 5G and beyond networks based on SDN and NFV*. *IEEE Network*, 34(3), 256-263.
- [3] Fourati, H., Maaloul, R., and Chaari, L. (2021). *A survey of 5G network systems: challenges and machine learning approaches*. *International Journal of Machine Learning and Cybernetics*, 12(2), 385-431.
- [4] Bega, D., Gramaglia, M., Fiore, M., Banchs, A., and Costa-Perez, X. (2019). *DeepCog: Optimizing resource provisioning in network slicing with AI-based capacity forecasting*. *IEEE Journal on Selected Areas in Communications*, 38(2), 361-376.
- [5] Buyakar, T. V. K., Agarwal, H., Tamma, B. R., and Franklin, A. A. *Resource allocation with admission control for GBR and delay QoS in 5G network slices*. In 2020 International Conference on Communication Systems and Networks (COMSNETS) (pp. 213-220). IEEE.
- [6] Prasanna, B., and Sai, B. (2023, November). *Dynamic Network Slice Orchestration using Gated Recurrent Units and Transformer-based Hybrid Model for IoT*. In 2023 9th International Conference on Smart Structures and Systems (ICSSS) (pp. 1-6). IEEE.
- [7] J.A. Villegas, C. Gijón, M. Toril, S. Luna Ramírez (2023). *Predicción del Rendimiento en Redes Celulares con Segmentación*. XXXVIII Simposio Nacional de la Unión Científica Internacional de Radio.
- [8] Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2022). *M5 accuracy competition: Results, findings, and conclusions*. *International Journal of Forecasting*, 38(4), 1346-1364.
- [9] Chen, S. A., Li, C. L., Yoder, N., Arik, S. O., and Pfister, T. (2023). *Tsmixer: An all-mlp architecture for time series forecasting*. arXiv preprint arXiv:2303.06053.
- [10] C. Gijón, M. Toril, S.Luna, *Data-Driven Estimation of Throughput Performance in Sliced Radio Access Networks via Supervised Learning*, *IEEE Transactions on Network and Service Management* (2022)
- [11] Hyndman, R. J., and Koehler, A. B. (2006). *Another look at measures of forecast accuracy*. *International journal of forecasting*, 22(4), 679-688.