ADVANCECRT

Centre for Research Training

Science Foundation Ireland sfi For what's next

Explaining Black-Box Models used in Resource Management for B5G/6G Networks

Pieter Barnard*, Irene Macaluso*, Nicola Marchetti*, Luiz A. DaSilva†
*{barnardp, macalusi, nicola.marchetti}@tcd.ie, †ldasilva@vt.edu

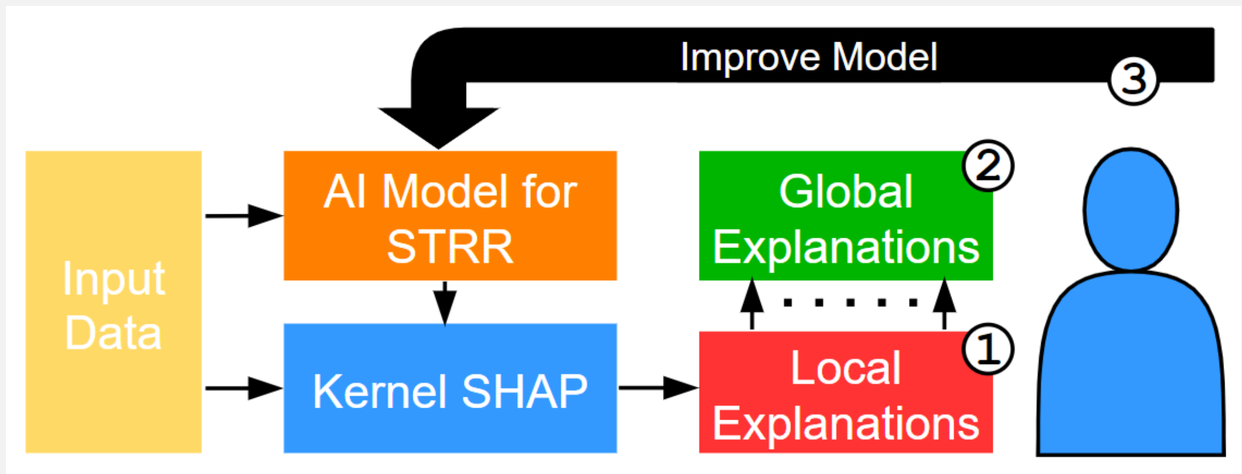
ABSTRACT: The growing complexity of 5G and next generation networks has sparked an upsurge in the use of AI within the networking domain. Although AI can achieve high accuracy in many contexts, AI models, such as deep neural networks lack transparency in their behaviour. This limits their adoption into high-stakes networking domains, such as resource management (RM), due to the potential threat posed to the reliable operation of the network. In this work, we present a framework based on state-of-the-art techniques from the field of Explainable AI (XAI) to help demystify the behaviour of complex models used in RM.

Introduction:

- AI is expected to become a key technology for network resource management (RM) in future beyond 5G (B5G) and 6G networks.
- However, the lack of transparency in AI models presents risks to the reliability and trustworthiness of the network.
- In short-term resource reservation (STRR), AI is used to make autonomous decisions about the network in real-time.
- STRR has implications on the revenue expenditure of the network operator (or slice tenant) as well as the quality of service (QoS) experienced by end users.

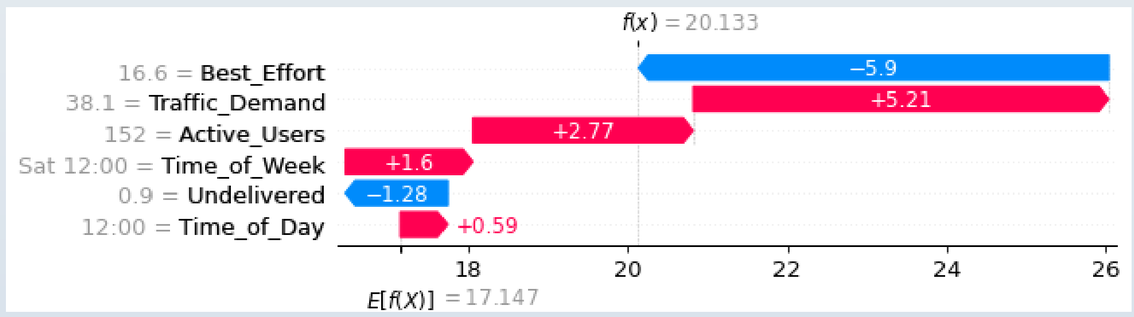
Proposed Solution:

- We present a framework based on SHAP values [1] from XAI to demystify black-box models used in RM, such as STRR.
- Our framework integrates XAI into the conventional AI pipeline and consists of three distinct stages**:**
  1. The use of local explanations to reveal how the model's features affect it's real-time decisions, i.e. explaining why a decision is made for a particular input.
  2. The use of global explanations to describe the general behaviour of the model, i.e. explaining how the model behaves across a wide input space.
  3. The use of explanations to debug potential flaws in the model before its final deployment.
- We apply our framework to the problem of STRR, in which a slice tenant uses AI to reserve resources from a network operator in a sliced 5G network.

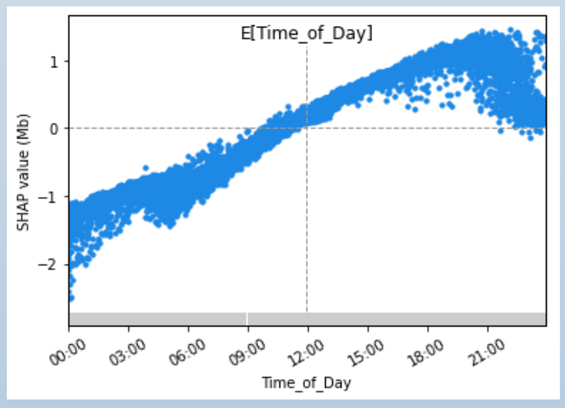


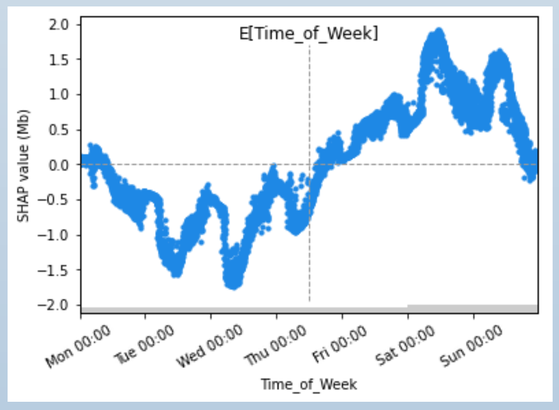
Results:

- A sample local explanation shows that best-effort traffic (related to the availability of network resources) intuitively reduces the reservation amount of the model, while high traffic demand increases it:



- A partial dependence plot (PDP), shows how the time of day affects the model's output globally. In this case, the PDP on the left corresponds to an early version of the model where an unintuitive 'jump' can be seen going from the end of the day to the start of the next. The PDP on the rights shows a re-trained model where the behaviour is more consistent with human intuition, engendering greater trust and reliability in the final model:





Conclusion:

- In future work, we hope to extend our framework to other networking problems, such as resource allocation, energy optimisation and self-healing etc. we also plan to extend our framework to include other forms of explanations, such as protypes and counterfactuals explanations etc.

References:

[1] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in neural information processing systems*, 2017, pp. 4765–4774.

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

Host Institution

UCC University College Cork, Ireland Coláiste na hOllscoile Corcaigh

CIT Cork Institute of Technology
Institiúid Teicneolaíochta Chorcaí

Maynooth University National University of Ireland Maynooth

TU Dublin Ollscoil Teicneolaíochta Bhaile Átha Cliath Technological University Dublin