USING MACHINE LEARNING FOR GREENING THE TRAIN TICKET APPLICATION

Carbon-aware autoscalers for microservices in the cloud

Abhinandan Roul, Ishara Madhavi, Mahira Joytu, Phuc Tran

EDISS · Åbo Akademi University · Università degli Studi dell'Aquila

INTRODUCTION

Cloud-based microservices are increasing energy demand in data centers, making sustainability a critical challenge in modern computing.

As digital infrastructure expands, we need smarter, more sustainable autoscaling solutions that optimize resource use and align with carbon-aware computing to reduce environmental impact without compromising performance.

OBJECTIVES

A smart autoscaling policy that improves resource utilization, optimizing for reduction in carbon emissions.

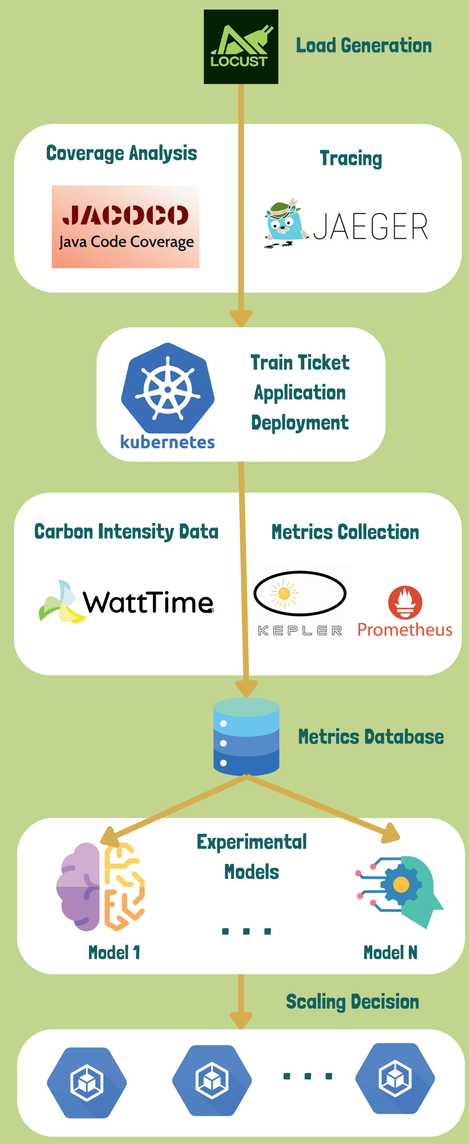
Energy Consumption
Inefficient scaling leads to higher energy usage

Carbon Intensity
No consideration about carbon footprint

Response Time
High response times impact user experience

To have a reasonable trade-off between sustainability and performance.

PIPELINE

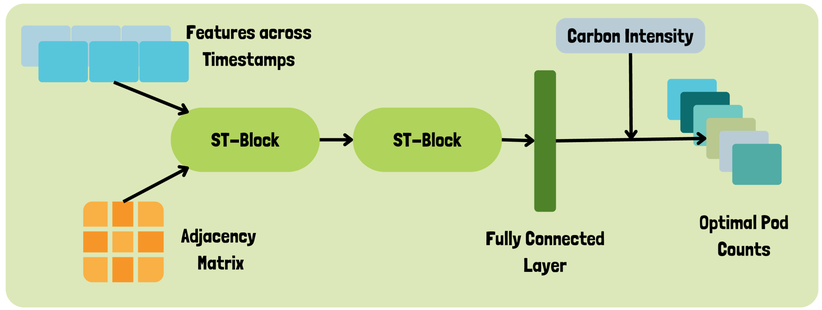


METHODOLOGY

MODELS

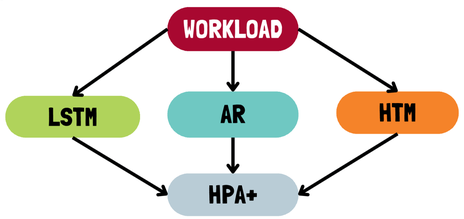
1 For Scaling Decision

Spatio-Temporal Graph Convolution Neural Network **(STGCN)**



2 For Workload Forecasting

Use timeseries forecasting models **(LSTM, AR, HTM)** to predict incoming web requests and make proactive scaling decisions.



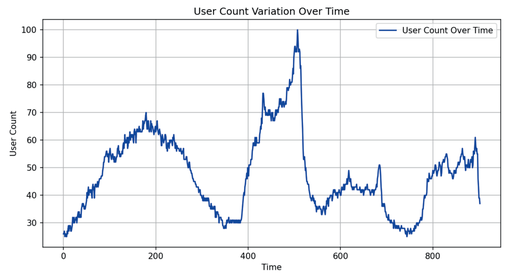
TRADE-OFF ANALYSIS

Formulate trade-off policies that prioritize either response time or carbon emissions, depending on situational demands.

ENERGY USAGE OF THE MODELS

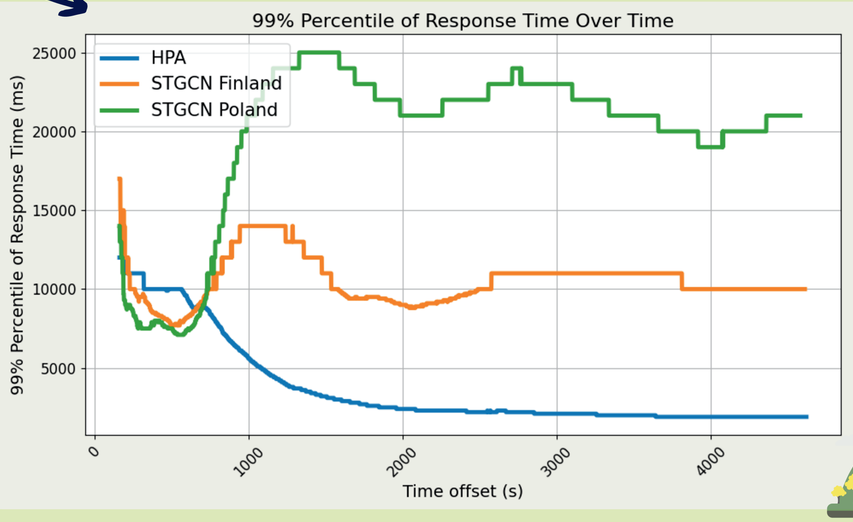
Analyze the energy overhead associated with both the training and inference processes.

DATASET



- **63 microservices** written in many languages: **Java, Go, Python**.
- Peak user count: **100 users**.
- Duration: **~ 3 hours**.
- Slight increases, slight decreases, sharp increases, sharp decreases, and fluctuations.

RESULTS



CARBON FOOTPRINT

Carbon Intensity		Finland (128 g CO2/kWh)	Poland (660 g CO2/kWh)
Carbon emissions	HPA	8.1 g	41.81 g
	STGCN	3.31 g	12.52 g

Energy usage varies depending on carbon intensity, scaling up when energy is cleaner and down when it's more carbon-intensive.

CHALLENGES

- Balancing sustainability and performance.
- Managing a complex microservices architecture.
- Handling workload variability.
- Ensuring accurate workload forecasting.
- Identifying useful metrics.