# ABSTRAK

*Fog computing* menjawab kebutuhan aplikasi *Internet of Things* (IoT) yang sensitif terhadap latensi dengan memperluas kapabilitas *cloud* ke jaringan. Tantangan utama dalam paradigma ini adalah *Fog Application Placement Problem* (FAPP), sebuah tugas optimasi multi-objektif yang kompleks dan *NP-hard*. Penempatan harus berhasil menyeimbangkan metrik performa yang saling bertentangan, seperti meminimalkan waktu respon aplikasi, konsumsi energi, dan biaya operasional, selagi mematuhi batasan sumber daya dari *fog nodes*. Meskipun algoritma metaheuristik *state-of-the-art* seperti WSGA dan NSGA-III dapat menghasilkan solusi berkualitas tinggi, sifatnya yang iteratif menyebabkan biaya komputasi yang tinggi dan waktu eksekusi yang lama, sehingga tidak praktis untuk lingkungan dinamis yang menuntut keputusan cepat. Penelitian ini mengusulkan dan memvalidasi kerangka kerja berbasis *learning* yang menyelesaikan *trade-off* kritis antara kecepatan dan kualitas. Kerangka kerja ini mengadaptasi model *Sequence-to-Sequence* (Seq2Seq), yang berfungsi bukan sebagai *optimizer* langsung, melainkan sebagai model prediktif cepat yang belajar untuk memetakan instans FAPP ke solusi penempatan berkualitas tinggi. Kebaruan pendekatan ini terletak pada dua mekanisme kunci: (1) metode kurasi data strategis yang melatih model pada *dataset* yang difilter menggunakan *utility function* yang bertujuan untuk mendekati sebuah *Pareto front* dari penempatan yang dihasilkan oleh metode heuristik, dan (2) mekanisme inferensi yang sadar-kendala yang mengintegrasikan validasi sumber daya *real-time* ke dalam proses *decoding* untuk memastikan semua penempatan yang dihasilkan *feasible*. Evaluasi empiris menunjukkan bahwa kerangka kerja *seq2seq* yang diusulkan mencapai kecepatan pengambilan keputusan yang sebanding dengan heuristik sederhana. Peningkatan kecepatan ini tidak mengorbankan kualitas solusi, ditunjukan dengan perbandingan dengan *benchmark* NSGA-III dan WSGA menggunakan analisis *Hypervolume*, yang menunjukan bahwa model menghasilkan solusi yang membentuk *Pareto front* yang secara statistik sebanding hingga secara signifikan lebih unggul seiring meningkatnya kompleksitas masalah. Analisis *trade-off* menunjukan kinerja model usulan menunjukkan adanya pengurangan pada metrik performa sambil menyeimbangkan *trade-off* antar metrik yang saling bersaing. Selanjutnya, analisis *robustness* menunjukkan bahwa kinerja kerangka kerja ini tidak sensitif terhadap urutan permintaan aplikasi maupun kompleksitas modul aplikasi. Temuan ini menetapkan bahwa kerangka kerja *data-driven* yang dilatih secara strategis ini menawarkan alternatif yang *robust*, *scalable*, dan efektif terhadap optimasi tradisional, yang memungkinkan penggunaannya sebagai manajemen sumber daya yang efisien di lingkungan *fog computing* yang dinamis.

***Kata kunci—****Fog Computing, Cloud Computing, Application Placement, Multi-criterion optimization and decision-making, Sequence-to-Sequence Models, Natural language processing, Supervised learning.*

# ABSTRACT

Fog computing addresses the needs of latency-sensitive Internet of Things (IoT) applications by extending cloud capabilities to the network. The central challenge within this paradigm is the Fog Application Placement Problem (FAPP), a complex, NP-hard multi-objective optimization task. A successful placement must balance conflicting performance metrics, such as minimizing application response time, energy consumption, and operational cost, while adhering to the resource constraints of fog nodes. While state-of-the-art metaheuristic algorithms like WSGA and NSGA-III can produce high-quality solutions, their iterative nature leads to high computational costs and long execution times, rendering them impractical for dynamic environments that demand rapid decisions. This research proposes and validates a learning-based framework that resolves this critical trade-off between speed and quality. The framework adapts a Sequence-to-Sequence (Seq2Seq) model, functioning not as a direct optimizer, but as a fast predictive model that learns to map FAPP instances to high-quality placement solutions. The novelty of this approach lies in two key mechanisms: (1) a strategic data curation method that trains the model on a filtered dataset using a utility function that aims to approximate a Pareto front of placements generated by heuristic methods, and (2) a constraint-aware inference mechanism that integrates real-time resource validation into the decoding process to ensure all generated placements are feasible. Empirical evaluation shows that the proposed seq2seq framework achieves decision-making speeds comparable to simple heuristics. This speed improvement does not compromise solution quality, as demonstrated by comparison with NSGA-III and WSGA benchmarks using Hypervolume analysis, which shows that the model produces solutions that form a Pareto front that is statistically comparable to significantly superior as problem complexity increases. Trade-off analysis shows the proposed model's performance demonstrates a reduction in performance metrics while balancing trade-offs between competing metrics. Furthermore, robustness analyses demonstrate that the framework's performance is insensitive to both the sequence of application requests and the complexity of application modules. The findings establish that this strategically trained, data-driven framework offers a robust, scalable, and effective alternative to traditional optimization, enabling its usage as efficient resource management in dynamic fog computing environments.

**Keywords—**Fog Computing, Cloud Computing, Application Placement, Multi-criterion optimization and decision-making, Sequence-to-Sequence Models, Natural language processing, Supervised learning.