

(12) PATENT APPLICATION PUBLICATION

(21) Application No.202511044783 A

(19) INDIA

(22) Date of filing of Application :08/05/2025

(43) Publication Date : 23/05/2025

(54) Title of the invention : AN ADAPTIVE SERVER LOAD PREDICTION AND AUTO-SCALING SYSTEM USING DEEP LEARNING

(51) International classification :G06Q0050060000, G06N0003080000, G06F0009500000, G06Q0010040000, H04L0067510000

(86) International Application No :NA
Filing Date :NA

(87) International Publication No : NA

(61) Patent of Addition to Application Number :NA
Filing Date :NA

(62) Divisional to Application Number :NA
Filing Date :NA

(71)Name of Applicant :

1)NOIDA INSTITUTE OF ENGINEERING & TECHNOLOGY

Address of Applicant :19, Knowledge Park-II, Institutional Area, Greater Noida – 201306, Uttar Pradesh, India. -----

Name of Applicant : NA

Address of Applicant : NA

(72)Name of Inventor :

1)KAKUMANU PRABHANJAN KUMAR

Address of Applicant :Department of Information Technology, Noida Institute of Engineering & Technology, Greater Noida. Greater Noida -----

(57) Abstract :

The present invention relates to an adaptive server load prediction and auto-scaling system utilizing deep learning. It comprises a data collection module (101), preprocessing unit (102), deep learning prediction engine (103), scaling decision module (104), resource management interface (105), system monitoring interface (106), and feedback loop mechanism (107). The system forecasts future server load using LSTM models and proactively scales resources in real-time. Continuous learning ensures high prediction accuracy, reduced latency, and efficient resource utilization, making it suitable for dynamic and scalable cloud computing environments.

No. of Pages : 16 No. of Claims : 5