CHENNAI UNIFIED METROPOLITAN TRANSPORT AUTHORITY
_____
TERMS OF REFERENCE

I. Introduction

At present the management of traffic and transportation system in Chennai Metropolitan Area (CMA) is fragmented and several government agencies are involved in it. For optimum utilization of resources, a greater coordination and streamlining of activities among all the concerned agencies is essential. Therefore, Government of India (GoI) insisted to establish Unified Metropolitan Transport Authority (UMTA) under its National Urban Transport Policy (NUTP) 2006 that was revised in 2014. Consequently, the Chennai Unified Metropolitan Transport Authority (CUMTA) was established by the Government of Tamil Nadu (GoTN) through the CUMTA Act, 2010 and the Rules for CUMTA were notified during January 2019.

The creation of CUMTA was a step towards integrated and sustainable urban mobility for Chennai. It is better placed to manage the existing institutional gaps and overlaps amongst various agencies in tackling urban transport issues at optimum utilization of resources. The jurisdiction of CUMTA extends over the whole of Chennai Metropolitan Area (CMA).

Currently, CUMTA is functioning under Housing and Urban Development Department with 'Special Officer' to execute the duties of Member-Secretary, CUMTA.

To fully operationalize CUMTA's functions in the urban mobility sector, additional experts/professionals, are now required. The role, key responsibilities, qualifications, work experience and additional competencies required for the post of Junior Data Scientist are given below:

II. Name of the Position: Junior Data Scientist
III. Description:

The Junior Data Scientist will support the team in implementing data analytics and developing ML / DL models for the CUMTA, which is a government transport nodal agency. The primary responsibility will be to brainstorm, conceptualize and develop solutions through Statistics, Machine learning or Deep learning in solving problems identified by the agency.

About Digital Chennai Project:

Chennai has witnessed rapid urbanization over the past decade, straining existing urban systems that are ill-suited to meet the scale or pace of emerging physical, economic and demographic demands. The compound challenges faced by public officials require new approaches to govern the urban ecosystem, and new tools to plan, implement, manage and monitor all the services.

The goal of this project, is to define the means and methods by which data will be compiled and integrated to inform planning and investment decisions, guide real-time urban management and operations, and allow for in-depth monitoring and evaluation of service delivery standards and other key performance indicators. Combining data from different agencies and carefully chosen external sources, in one platform should provide a multi-sectoral, holistic understanding of city performance and the ability to make empirically informed choices between competing policies and programs.

Against this backdrop, the two technology solutions envisaged are as follows. The **Integrated Urban Data Platform (IUDP)** will comprise a platform that is able to integrate, store and share spatial and non-spatial (both structured and non-structured) data through a single data portal; and provide modular analytics that combine and analyze the data for a comprehensive tracking of city service performance. The **Urban Project Planning and Management (UPPM)** tool will facilitate planning, management, coordination, supervision, and monitoring of multi-agency urban projects (capital and operations).

IV. **Key Responsibilities:**
- Research and test novel machine learning approaches by analyzing datasets from multiple sources.
- Understanding and designing data schemas/stores and data pipelines for storing and processing of different kind of structured and unstructured data sets like tables, image and video feeds, process data, email data, social feed data and many more
- Connecting to Data sources, importing data, and transforming data for Required Processing/ Reporting/ Business Intelligence
- Develop powerful data stories and visualization in the form of Dashboards using Tableau or PowerBI

- Working closely with our backend engineering team to build a robust suite of libraries for extracting and manipulating data for the Digital Chennai Platform.
- Develop production-ready implementations of proposed solutions across different ML and DL algorithms, including testing on live data to improve accuracy, efficacy, and robustness
- Create use cases with domain experts for solving a business problem.
- Assist in developing and implementing data integration and management solutions that meet the requirements of the department.
- Collaborate with other team members to ensure that data management and integration activities are coordinated and aligned with agency objectives.
- Support data collection and data creation efforts for CUMTA

V. **Qualifications and Experience:**

- Bachelor's degree in Computer Science, Information technology, Mathematics, Statistics, Data science, ML/AI or a related field is required. A master's degree in a relevant field is preferred but not mandatory. Certification courses in Big-Data/Cloud platforms/Data Science/ML/GIS will be an added advantage.
- Minimum **3 years** of experience in statistics, data science, data integration, and analysis, preferably in a large-scale, complex environment preferably in transport sector.
- In-depth working, beyond coursework, familiarity with statistical techniques and current ML techniques, both supervised and unsupervised learning techniques and other ML Techniques.
- Experience in working on modelling spatiotemporal systems such as GIS, urban dataset, satellite images, etc.
- Implementation experiences and deep knowledge of Classification, Time Series Analysis, Pattern Recognition, Reinforcement Learning, Deep Learning, Dynamic Programming and Optimization.
- Experience in developing and deploying on one or more Cloud Environments (GCP/Azure/ AWS). Familiarity with Flask / Django Framework.
- Experience on end-to-end ML models deployment in production (MLOps) in cloud or on-prem.
- Knowledge of ETL tools, data cleaning and preprocessing pipeline, ML training and ML model tracking frameworks such as Kubeflow, MLFlow, etc.
- Familiarity with well-known Python packages like Pandas, Numpy, etc., and DL frameworks Keras, TensorFlow, PyTorch. And knowledge of Big Data tools and environment.

- Proficiency in SQL, database technologies such as MySQL/No SQL is an added advantage.
- Fundamental knowledge of distributed data processing such as Spark and streaming platforms such as Kafka.
- Programming skills in Python is a mandatory. Good critical and algorithmic thinking, verbal and written communication skills.

VI. **Work Location**

The work location will be at Chennai Unified Metropolitan Transport Authority, Nandanam, Chennai. She / He will be required to travel frequently within and outside the city and occasionally outside the state for project purposes.

VII. **Duration**

Appointment to the post will be on contract basis initially for a period of two year and likely to be extended based on the performance.

VIII. **Reporting Arrangements**

The Junior Data Scientist will report to the Special officer / Member Secretary, CUMTA.

IX. **Remuneration**

Upto Rs.80,000 per month. Remuneration will not be a constraint for exceptionally deserving candidates.

X. **Language**

Working knowledge in Tamil is desirable.

Apply by:

Interested candidates may apply with the detailed resume, portfolio, expected salary and pay-slip for the last 3 months on or before **10-August-2023** by e-mail to cumtaoffice@tn.gov.in.