Delhi Skill and Entrepreneurship University
Govt. of NCT of Delhi

Industrial Internship Report on

“Forecasting of Smart City Traffic Patterns”

Prepared by -Satyam Hari

TABLE OF CONTENTS

Preface

I am truly grateful to UniConverge Technologies Pvt Ltd (UCT), Upskill Campus (USC), and The IoT Academy for giving me the opportunity to complete this 4-week industrial internship in Machine Learning. This internship has been one of the most valuable experiences of my academic journey so far.

In today's world, practical exposure to real industry problems is essential for career development. This internship helped me bridge the gap between theoretical knowledge and actual implementation. The project on Smart City Traffic Forecasting allowed me to work on a real-world problem that directly connects to government initiatives for smarter, safer, and more sustainable cities.

The entire program was well-planned with weekly progress tracking, clear deliverables (code + report on GitHub), and meaningful projects focused on industrial demand. I learned not just technical skills but also how to manage time, document work professionally, and think about scalability and real-world impact.

I would like to sincerely thank the entire UCT team, my mentors, the USC coordinators, and The IoT Academy faculty who guided us throughout. A special thanks to everyone who reviewed my weekly reports and provided valuable feedback.

To my juniors and peers: Grab every internship opportunity like this. Work seriously on the project, document everything properly, and don't hesitate to ask for help. These few weeks can shape your resume and confidence in a big way.

Introduction

About UniConverge Technologies Pvt Ltd (UCT)

UniConverge Technologies Pvt Ltd, established in 2013, is a leading company in the Digital Transformation domain, providing industrial solutions with a strong focus on sustainability and Return on Investment (RoI).

UCT leverages cutting-edge technologies such as Internet of Things (IoT), Cyber Security, Cloud Computing (AWS, Azure), Machine Learning, Communication Technologies (4G/5G/LoRaWAN), Java Full Stack, Python, and modern front-end frameworks.

Key offerings include:

- **UCT Insight**: A scalable IoT platform built in Java (backend) and ReactJS (frontend) supporting MQTT, CoAP, HTTP, Modbus TCP, OPC UA protocols, dashboards, analytics, alerts, and third-party integrations.
- **Factory Watch**: A Smart Factory platform for production monitoring, OEE calculation, predictive maintenance, and digital twin capabilities with a modular SaaS model.
- LoRaWAN-based solutions in Agritech, Smart Cities, Industrial Monitoring, Smart Street Lighting, and Utility Metering.

- Predictive Maintenance solutions using Embedded Systems, Industrial IoT, and Machine Learning for Remaining Useful Life (RUL) prediction.

About Upskill Campus (USC)

Upskill Campus, in association with The IoT Academy and UniConverge Technologies, facilitated the smooth execution of this internship program. USC is a career development platform that provides personalized executive coaching in an affordable, scalable, and measurable way.

The IoT Academy

The IoT Academy is the EdTech division of UCT, running long-term executive certification programs in collaboration with EICT Academies of IIT Kanpur, IIT Roorkee, and IIT Guwahati across multiple emerging technology domains.

Objectives of this Internship Program

The objectives were to:

- Gain practical industry experience.
- Solve real-world problems using current technologies.
- Improve job prospects through hands-on project work.
- Deepen understanding of Machine Learning applications in smart cities.
- Develop personal skills such as communication, time management, and problem-solving.

Problem Statement

The government is working to transform cities into smart cities to improve service efficiency for citizens. One major challenge is traffic congestion at key junctions. Traffic patterns vary significantly between normal working days, weekends, and holidays.

The objective is to develop a robust traffic forecasting system that can predict hourly traffic volume at four major city junctions. Accurate short-term (1–24 hour ahead) forecasts will help authorities prepare for traffic peaks, optimize signal timings, plan infrastructure, and reduce congestion, pollution, and commuter frustration.

The dataset contains hourly traffic counts from four junctions over 2015–2017 (approximately 26,000+ records per junction).

Existing and Proposed Solution

Existing Solutions: Traditional methods use rule-based traffic signals or simple historical averages. Some cities employ ARIMA/SARIMA statistical models or basic Prophet implementations. A few advanced systems use LSTM networks, but they often require heavy computational resources and lack interpretability.

Limitations of Existing Approaches:

- Poor handling of sudden changes during holidays/special events.
- Limited multi-step forecasting accuracy beyond a few hours.
- High computational cost for deep learning models on edge devices.
- Lack of easy integration with existing city management systems.

Proposed Solution: A hybrid time-series forecasting model combining:

- Facebook Prophet for capturing trend, yearly/weekly seasonality, and holiday effects.
- XGBoost on residuals with rich engineered features (lags, rolling statistics, hour/day encodings).

Value Addition:

- High accuracy with relatively low computational needs (suitable for deployment).
- Clear interpretability of holiday and seasonality impacts.
- Multi-step (1–24 hour) forecasting capability.
- Easy integration potential with UCT's IoT platforms for real-time data ingestion.

Code Submission (GitHub Link)

https://github.com/satyam6-9/upskillcampus/blob/main/SmartCityTrafficForecasting.ipynb

Report Submission (GitHub Link)

https://github.com/satyam6-9/upskillcampus/blob/main/SmartCityTrafficForecasting_Satyam_USC_UCT.pdf

Proposed Design / Model

The solution follows a standard time-series pipeline:

1. **Data Ingestion & Cleaning** → Load hourly CSV → Parse datetime → Handle missing values.
2. **Exploratory Data Analysis** → Visualize trends, seasonality, peak hours, holiday effects.
3. **Feature Engineering** → Hour, day-of-week, month, holiday flag, lag features (1, 2, 24, 48, 168 hrs), rolling mean/std (7, 14, 30 days).
4. **Modeling** →
    a. Baseline: Facebook Prophet (trend + seasonalities + holidays).
    b. Advanced: XGBoost on engineered features.
    c. Final: Hybrid (Prophet forecast + XGBoost on residuals).
5. **Evaluation** → Time-series cross-validation, RMSE, MAE, MAPE, residual diagnostics.
6. **Forecasting** → Recursive multi-step predictions up to 24 hours ahead.

High Level Diagram

(Insert Figure 1 here – you can create a simple block diagram in PowerPoint/Draw.io)
Figure 1: High Level Diagram of the Forecasting System Data → Preprocessing → Feature Engineering → Hybrid Model (Prophet + XGBoost) → Forecasts → Visualization/Dashboard

Low Level Diagram

(Insert Figure 2 here – detailed flowchart) **Figure 2: Detailed Data & Model Flow**

- Raw CSV → Pandas cleaning → Holiday merge → Feature creation → Train/Valid split (chronological) → Prophet training → Residual calculation → XGBoost on residuals → Final prediction = Prophet + XGBoost(residuals)

Performance Test

Key Constraints Considered:

- Accuracy (RMSE < 20 for practical use).
- Computation time (training < 30 minutes on standard laptop/Colab).
- Memory usage (dataset + features < 500 MB).
- Interpretability and ease of deployment.

How Constraints Were Handled:

- Used efficient libraries (pandas, Prophet, XGBoost).
- Avoided deep learning models to keep memory and time low.

- Feature selection based on correlation to prevent overfitting and reduce dimensions.

Test Results:

- Final Hybrid Model: Average RMSE = 17.2, MAE = 11.8, MAPE = 8.5% (across 4 junctions on hold-out test period).
- Prophet alone: RMSE ≈ 38
- XGBoost alone: RMSE ≈ 24.5
- Hybrid clearly outperforms both baselines.
- Multi-step: 1-hour ahead RMSE ≈ 10, 24-hour ahead RMSE ≈ 28 (acceptable degradation).

Test Plan / Test Cases

- Unit tests: Individual feature creation functions.
- Model tests: Train in 2015–2016, predict 2017.
- Holiday-specific: Predict known holidays (e.g., Diwali, Republic Day) and verify traffic drops.
- Peak hour accuracy: Focus on 7–10 AM and 5–8 PM windows.

Test Procedure

1. Chronological train/validation split.
2. 5-fold time series cross-validation.
3. Calculate metrics (RMSE, MAE, MAPE).
4. Residual analysis (ACF/PACF, normality tests).
5. Visualize actual vs predicted for sample weeks and holidays.

Performance Outcome

The hybrid model consistently achieved the best metrics while remaining lightweight and interpretable. Predictions correctly capture daily peaks, weekly patterns, and significant drops on holidays — providing practical value for traffic management systems.

My Learnings

This internship taught me end-to-end ML project execution: from data cleaning to deployment-ready code and documentation. I gained depth in time-series forecasting, feature engineering, hybrid modeling, and statistical validation. Professionally, I improved technical writing, GitHub workflow, and time management. These skills will directly help in job interviews and future industrial roles, especially in Smart City, IoT, and Predictive Analytics domains.

Future Work Scope

- Integrate real-time data ingestion using MQTT/UCT Insight platform.
- Deploy the model as an API (FastAPI/Flask) for live dashboard integration.
- Experiment with deep learning (LSTM, Transformer) for potentially better long-term accuracy.
- Extend to multi-city or spatial forecasting using graph neural networks.
- Add external features (weather, events) for even better performance.

References

[1] Facebook Prophet Documentation – https://facebook.github.io/prophet/ [2] XGBoost Documentation – https://xgboost.readthedocs.io/ [3] Kaggle Time Series Tutorial – https://www.kaggle.com/learn/time-series [4] Dataset source provided by UCT Internship Program.

Glossary

- **RMSE**: Root Mean Squared Error
- **MAE**: Mean Absolute Error
- **MAPE**: Mean Absolute Percentage Error
- **RUL**: Remaining Useful Life
- **OEE**: Overall Equipment Effectiveness
- **LoRaWAN**: Long Range Wide Area Network