

CROWDCAST: PREDICTING MALL VISITORS OVER TIME

DISSERTATION

Submitted partial fulfillment of the requirements of the
MTech in Artificial Intelligence & Machine Learning

By

HANUMANTHA REDDY KATREDDY
(2023AB05024)

Under the supervision of

SURESH BABU BALLA
Project Director

Dissertation work carried out at
Blue Yonder India Pvt Ltd, Bangalore

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE
Pilani (Rajasthan) INDIA

MAY 2025

BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI
SECOND SEMESTER 2024-25

DSECLZG628T / AIMLCZG628T DISSERTATION

Dissertation Title: CROWDCAST: PREDICTING MALL VISITORS OVER TIME

Name of Supervisor: SURESH BABU BALLA

Name of Student: HANUMANTHA REDDY KATREDDY

ID No. of Student: 2023AB05024

Courses Relevant for the Project & Corresponding Semester

Artificial Intelligence & Machine Learning

1. Time Series Analysis and Forecasting
2. Machine Learning for Predictive Modeling
3. Data Mining and Pattern Recognition
4. Deep Learning (especially RNNs, LSTMs for sequential data)

Mathematics & Statistics

1. Probability and Statistics
2. Linear Algebra and Calculus for Data Science
3. Statistical Inference and Hypothesis Testing

Programming & Tools

1. Python for Data Science
2. Data Visualization with Python
3. Data Wrangling and Preprocessing

Abstract

The primary objective of this project is to forecast mall footfall at any given future time using historical and contextual data. The analysis leverages two key data sources: (1) hourly entry counts recorded at various mall entrances (referred to as Gates), and (2) supplementary contextual data such as weather conditions and public holidays.

In addition to accurate prediction, the project aims to develop a deep understanding of the underlying data patterns and extract actionable business insights. These insights will support more efficient forecasting and inform strategic decision-making for mall operations and resource planning.

Key Words:

- Mall Footfall Prediction
- Time Series Forecasting
- Retail Analytics
- Visitor Flow Modeling
- Predictive Analytics
- Business Intelligence
- Weather Impact on Retail
- Holiday Effect on Footfall
- Data-Driven Decision Making
- Gate-wise Entry Analysis
- Temporal Data Analysis
- Smart Retail Solutions
- Customer Traffic Forecasting
- Data Visualization
- Operational Planning

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI
II SEMESTER 24-25
DSECLZG628T / AIMLCZG628T DISSERTATION
Dissertation Outline (Abstract)

BITS ID No.2023ab05024

Name of Student: Hanumantha Katreddy

Name of Supervisor: Suresh Babu Balla

| | Supervisor | Student |
|--------------------------------|--|--|
| Name | Suresh Babu Balla | Hanumantha Reddy Katreddy |
| Qualification | B.Tech | M.Sc.IS |
| Designation and address | Project Director | Sr. QA Architect |
| Email Address | suresh.balla@blueyonder.com | hanumantha.katreddy@blueyonder.com |
| Topic of Dissertation | CROWDCAST: PREDICTING MALL VISITORS OVER TIME | |



HANUMANTHA KATREDDY
(Signature of Student)

Date: 23/05/2025



SURESH BABU BALLA
(Signature of Supervisor)

Date: 25/05/2025

Project Work Title: CROWDCAST: PREDICTING MALL VISITORS OVER TIME

1. Purpose of the Work and Expected Outcome

The primary purpose of this project is to develop a predictive model that can accurately forecast mall footfall at various entry points (Gates) for any given future time. By leveraging historical footfall data along with contextual variables such as weather and holidays, the project aims to support mall management in making data-driven decisions. The expected outcomes include:

- A robust time series forecasting model.
- Actionable business insights derived from data analysis.
- Improved operational planning and resource allocation.

2. Literature Review

- Several studies have explored footfall prediction in retail environments using time series and machine learning techniques:
- Zhou et al. (2020) compared SARIMAX, Prophet, XGBoost, and LSTM for hourly footfall forecasting, highlighting the importance of external variables like weather and holidays.
- Murcio and Wang (2025) evaluated pedestrian flow forecasting methods using Wi-Fi sensor data, emphasizing the effectiveness of LSTM models.
- Other works have focused on the impact of footfall on mall performance, customer behavior, and retail strategy.
- These studies provide a strong foundation for selecting appropriate models and understanding the influence of external factors on footfall patterns.

3. Existing Process and Its Limitations

- Traditionally, mall footfall estimation relies on manual counting, basic sensors, or simple statistical averages. These methods have several limitations:
- Lack of accuracy: They fail to capture complex temporal patterns and external influences.
- No predictive capability: They are reactive rather than proactive.
- Limited scalability: Manual methods are not feasible for large-scale or real-time forecasting.

4. Justification for Methodology Selection

- Time series forecasting methods such as LSTM are well-suited for modeling sequential data like hourly footfall. The inclusion of exogenous variables (weather, holidays) through models like hybrid ML models (e.g., XGBoost with time features) enhances predictive accuracy. LSTM networks are particularly effective for capturing long-term dependencies in time series data.

5. Project Work Methodology

The project follows these key steps:

- Data Collection: Gather hourly footfall data, weather data, and holiday calendars.
- Data Preprocessing: Handle missing values, normalize data, and engineer time-based features.
- Exploratory Data Analysis (EDA): Visualize trends, seasonality, and correlations.

- Model Selection and Training: Compare models like SARIMA, Prophet, and LSTM.
- Model Evaluation: Use metrics such as RMSE, MAE, and MAPE to assess performance.
- Insight Generation: Analyze model outputs to derive business insights.
- Deployment (Optional): Create a dashboard or API for real-time predictions.

6. Benefits Derivable from Work

- Operational Efficiency: Better staff scheduling, energy management, and security planning.
- Marketing Optimization: Targeted promotions based on expected footfall.
- Customer Experience: Improved crowd management and service delivery.
- Strategic Planning: Data-driven decisions for store placement and event planning.

7. Additional Supporting Details

- Scalability: The model can be extended to multiple malls or integrated with IoT systems.
- Real-Time Adaptability: With live data feeds, the model can provide near real-time forecasts.
- Cross-Domain Application: The methodology is applicable to airports, stadiums, and public transport hubs.

1. Broad Area of Work

Machine Learning and Predictive Analytics in Retail Operations

2. Objectives

The objectives of my project are as follows:

- The objective of this project is to develop a predictive model that accurately forecasts mall footfall at various entry points (Gates) for any given future time.
- This will be achieved by analyzing historical hourly footfall data in conjunction with supplementary contextual data such as weather conditions and public holidays.
- The project also aims to extract meaningful business insights from the data to support strategic planning, optimize resource allocation, and enhance operational efficiency within the mall environment

3. Scope of Work

The scope of the work involves understanding the existing system more and other required documents.

- Gathering or exploring the data which is useful for the problem.
- Understanding the data nature to perform wrangling, exploration and cleaning.
- Model training and looking for best model for better performance.
- Making model available for usage.

Note: Since our company is not willing to share the data needed for the project, I will be working on the data which is available on the public domain.

4. Detailed Plan of Work (for 16 weeks)

| Sl. No. | Tasks or subtasks | Start Date-End Date | Planned duration in weeks | Specific Deliverable |
|---------|---|--------------------------|---------------------------|--|
| 1 | As part of Data gathering and Cleaning: Importing the Sales history data of different SKUs from our customers DB2 data base and joining multiple data sets , detect missing values (of daily/weekly sales) if any and check for anomalies, Imputing the missing values if required and reshaping the data | 12/05/2025 to 01/06/2025 | 3 | Data Cleansing and gathering the data |
| 2 | As part of Exploratory Data Analysis visualizing the different sales patterns of the SKUs and gaining more insights into the data by using matplotlib, Histograms. Try to get the insights into correlation other than seasonality | 02/06/2026 to 15/06/2025 | 2 | Insights into the different SKUs sales pattern and Visualization |
| 3 | Pre-Processing and Model Selection: pre-processing and splitting the data into train and test sets and selecting the RNN Model with Input, dropout and dense layers using Activation functions: ReLU, Sigmoid ,Tanh | 16/06/2025 to 13/07/2025 | 4 | Trained model |
| 4 | Developing and fine tuning hyperparameters of the RNN Model to minimize Mean Square error using different activation method of optimization. | 14/07/2025 to 27/07/2025 | 2 | Model with good accuracy |

| | | | | |
|---|--|--------------------------|---|--|
| 5 | Comparing the Forecast error and its accuracy with our products existing Forecasting Techniques. | 28/07/2025 to 10/08/2025 | 2 | Comparison of Forecast accuracy between the RNN Model and our product techniques |
|---|--|--------------------------|---|--|

5. Literature References

The following are referred journals from the preliminary literature review

1. **Planning and Forecasting Using Time Series Modelling:** <https://www.cdrc.ac.uk/wp-content/uploads/sites/72/2020/07/Yang-Zhou.pdf>
2. **Identify Optimal Pedestrian Flow Forecasting Methods in Great Britain Retail Areas:** <https://www.mdpi.com/2220-9964/14/2/50>
3. **A Study to Assess the Impact of Footfalls and Category Mix on Mall Performance:** https://www.academia.edu/36065557/A_STUDY_TO_ASSESS_THE_IMPACT_OF_FOOTFALLS_A_ND_CATEGORY_MIX_ON_MALL_PERFORMANCE

Supervisor's Rating of the Technical Quality of this Dissertation Outline

EXCELLENT / GOOD / FAIR/ POOR (Please specify): GOOD

Supervisor's suggestions and remarks about the outline (if applicable).

Date: 25/05/2025

B Suresh Babu

(Signature of Supervisor)

Name of the supervisor: SURESH BALLA

Email Id of Supervisor:

Mob # of supervisor: 9885710140