DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOLOGIES, NUZVID

TourFlow: Forecasting Tourist Visits and Recommending Nearby Attractions Using Machine Learning

by

N200556 (T. V S Likhitha Bhavani)
N200541 (L. Sravani Sindhu)
N201014 (A. Sasi Rekha)
N200329 (Kalyan Lokireddy)
N200808 (M.Girish)

Report submitted to

Rajiv Gandhi University of Knowledge Technologies, Nuzvid

for the fulfillment of Mini Project

Of

Bachelor of Technology
in
Computer Science and Engineering

RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOLOGIES
Department of Computer Science and Engineering
RGUKT-Nuzvid, Krishna District – 521202

Declaration

We hereby declare that:

1. The contents of this report are the result of our original work carried out under the guidance of our project supervisor.
2. This report has not been submitted to any other institution or university for the award of any degree, diploma, or certification.
3. We have strictly followed the project submission guidelines as prescribed by the Institute throughout the preparation of this report.
4. We affirm that the project work complies with the ethical standards and academic integrity policies set by the Institute.
5. All external sources of information, including data, theories, figures, and text, have been duly acknowledged and cited. Where applicable, necessary permissions were obtained from the copyright holders.

N200556	T. V S Likhitha Bhavani
N200541	L. Sravani Sindhu
N201014	A. Sasi Rekha
N200329	Kalyan Lokireddy
N200808	M.Girish

RAJIV GANDHI UNIVERSITY OF KNOWLEDGE TECHNOLOGIES
Department of Computer Science and Engineering
RGUKT-Nuzvid, Krishna District – 521202

CERTIFICATE

This is to certify that the mini-project report titled **"Forecasting Tourist Visits and Recommending Nearby Attractions Using Machine Learning"** submitted by **Ms. T. V S Likhitha Bhavani,Ms. L. Sravani Sindhu,Ms. A. Sasi Rekha,Mr. Kalyan Lokireddy,Mr. M.Girish** to Rajiv Gandhi University of Knowledge Technologies, Nuzvid, India, is a record of the bonafide project work carried out by the students under my supervision and guidance. This work is considered worthy of submission for the fulfillment of the Mini Project requirement for the award of the degree of **Bachelor of Technology in Computer Science and Engineering** at the Institute.

Dr. D.V Nagarjuna Devi mam
Project Guide
Faculty, Department of CSE
Rajiv Gandhi University of Knowledge Technologies, Nuzvid

Y. Kalavathi mam
project Examiner
Faculty, Department of CSE
Rajiv Gandhi University of Knowledge Technologies, Nuzvid

Acknowledgement

We would like to express our sincere gratitude to our project guide, Dr. D.V Nagarjuna Devi, for her exceptional guidance, continuous support, and valuable insights throughout the course of this project. His expertise and encouragement have been instrumental in the successful completion of our work, and we deeply appreciate the opportunity to learn under his mentorship.

We are also thankful to the Head of the Department, CSE, and all the faculty members of Rajiv Gandhi University of Knowledge Technologies, Nuzvid, for their constant motivation and support, which provided a conducive environment for research and learning.

We extend our heartfelt thanks to our families and friends for their unwavering support, blessings, and encouragement during every stage of this project.

N200556	**T. V S Likhitha Bhavani**
N200541	**L. Sravani Sindhu**
N201014	**A. Sasi Rekha**
N200329	**Kalyan Lokireddy**
N200808	**M.Girish**

Abstract

Tourism plays a critical role in regional economic development, requiring effective planning and resource management to ensure a positive visitor experience. This project, titled **"Forecasting Tourist Visits and Recommending Nearby Attractions using Machine Learning"**, presents a data-driven framework to predict daily tourist footfall and provide location-aware recommendations to visitors.

To accurately forecast tourist volume, we engineered temporal and event-based features including **date of visit**, **weekend indicators**, **holiday flags**, and **festival impact scores**, the latter derived using an exponential smoothing technique on historical attendance data. Among several machine learning models evaluated, the **Extra Trees Regressor (ETR)** demonstrated the best performance, achieving a predictive accuracy of **86%**. ETR was selected over alternatives such as Random Forest due to its ability to reduce both variance and overfitting through the introduction of greater randomness in feature splits. Moreover, ETR is computationally efficient and well-suited for handling high-dimensional and non-linear relationships commonly present in time-series forecasting tasks.

For the secondary objective of recommending nearby tourist attractions, we implemented a **proximity-based recommendation engine** using the **cKDTree algorithm from the Nearest Neighbors class**. The cKDTree structure enables fast spatial queries, making it highly effective for real-time use cases. A distinguishing advantage of this method is its **logarithmic query time complexity** and **low memory footprint**, which allows for scalable deployment even in resource-constrained environments.

The integrated system not only assists local authorities in proactive decision-making but also enhances the tourist experience by dynamically suggesting nearby places of interest. This combination of predictive analytics and intelligent recommendation offers a practical and scalable solution for smart tourism initiatives.

Table of Contents

List of Figures

Introduction

Tourism is a vital contributor to economic development across the globe. With its ability to generate employment, stimulate regional economies, and support infrastructure development, the tourism industry plays a significant role in national and state-level GDP growth. As the sector continues to expand, forecasting tourist volume has become increasingly important for policymakers, tourism boards, and local businesses. Accurate tourism demand forecasting provides essential data for planning, resource allocation, and the development of strategic marketing initiatives.

This project, titled **"Forecasting Tourist Visits and Recommending Nearby Attractions,"** focuses on the renowned Tirumala Venkateswara Temple—a prominent pilgrimage destination located in Tirumala near Tirupati, Andhra Pradesh. As one of the most visited religious sites in India, it draws millions of devotees annually. The temple is managed by Tirumala Tirupati Devasthanams (TTD), which proposed a budget of 3,096.40 crore for the year 2022–2023, marking an increase of 158.58 crore from the previous year. This upward trend underscores the growing significance of pilgrimage tourism in Andhra Pradesh and its substantial contribution to the state's economy.

In this digital age, tourists increasingly rely on online platforms for travel research and planning. Search engines, in particular, offer valuable insights into tourist behavior and interest. By analyzing relevant search keywords—such as *Tirupati*, *Tirumala*, *VIP darshan*, and *Tirumala darshan*—we can anticipate fluctuations in visitor interest and predict tourist arrivals more effectively.

This project leverages search trend data, historical footfall patterns, and seasonal variations to forecast future tourist visits to Tirumala. In addition, it aims to recommend nearby attractions to encourage extended stays and improve the overall travel experience. Key nearby sites such as *Sri Padmavathi Ammavari Temple*, *Srikalahasti Temple*, *Kapila Theertham*, and *Talakona Waterfalls* will be included in the recommendations, offering a holistic view of the region's tourism potential.

By combining data-driven forecasting techniques with intelligent recommendation systems, this project seeks to support tourism planning, enhance visitor experience, and strengthen the economic impact of tourism in the region.

Machine learning algorithms play a key role in identifying patterns in tourism arrivals and forecasting future trends. With growing demand for tourist transportation, especially during peak seasons, pressure on infrastructure can lead to overcrowding and safety risks.

Accurate tourism demand forecasting enables efficient resource allocation and helps improve both the safety and effectiveness of travel operations.

Background and Related Work

The tourism industry has been the subject of numerous studies, highlighting its economic significance in various regions. In this context, predicting tourist arrivals has emerged as a crucial area of research. This study focuses on forecasting the number of pilgrim arrivals at the Tirumala Tirupati Devasthanams (TTD), a major pilgrimage site, by utilizing regression models that incorporate highly correlated features such as weekends, holidays, and festival names.

A range of machine learning models have been explored for tourism forecasting and place recommendation tasks. While many studies rely on similar datasets, the models and objectives often differ based on the geographical focus and specific aims of the research. In particular, various algorithms have been employed to predict daily tourism volumes for attractions, with each model tailored to the nature of the data and the forecasting challenge.

The Extra Trees Regressor has been proposed as an effective tool for predicting the number of pilgrims visiting TTD, leveraging input variables such as historical data, search engine trends, and weather conditions. This combination of features enables the model to make highly accurate forecasts, demonstrating the suitability of the Extra Trees Regressor for this task.

Additionally, for recommending nearby attractions, the CKD (Centroid-Kernel Distance) algorithm is employed to identify and suggest the most relevant places of interest in proximity to TTD. This algorithm has proven effective in providing personalized recommendations based on spatial relationships and visitor preferences, further enhancing the tourism experience.

Methodology

Proposed Methodology for the Tourist Forecasting:

The proposed approach for forecasting tourist arrivals using the **Extra Trees Regressor model** is outlined in fig1. The methodology consists of several key stages: data collection, preprocessing, feature selection, model training, validation, and prediction.

The process begins with data collection and preparation, which includes gathering historical visitor data, weather information, holidays, and search engine trends. Following this, Pearson correlation analysis is conducted to select the most relevant features—specifically, those with a correlation coefficient greater than 0.2 or less than -0.2.

Subsequently, the selected features are used to train the Extra Trees Regressor model. This ensemble learning technique is particularly effective due to its ability to handle high-dimensional data and capture complex, non-linear relationships. The model is then validated using appropriate metrics to assess performance.

The final phase predicts tourist arrivals to support planning and decision-making.

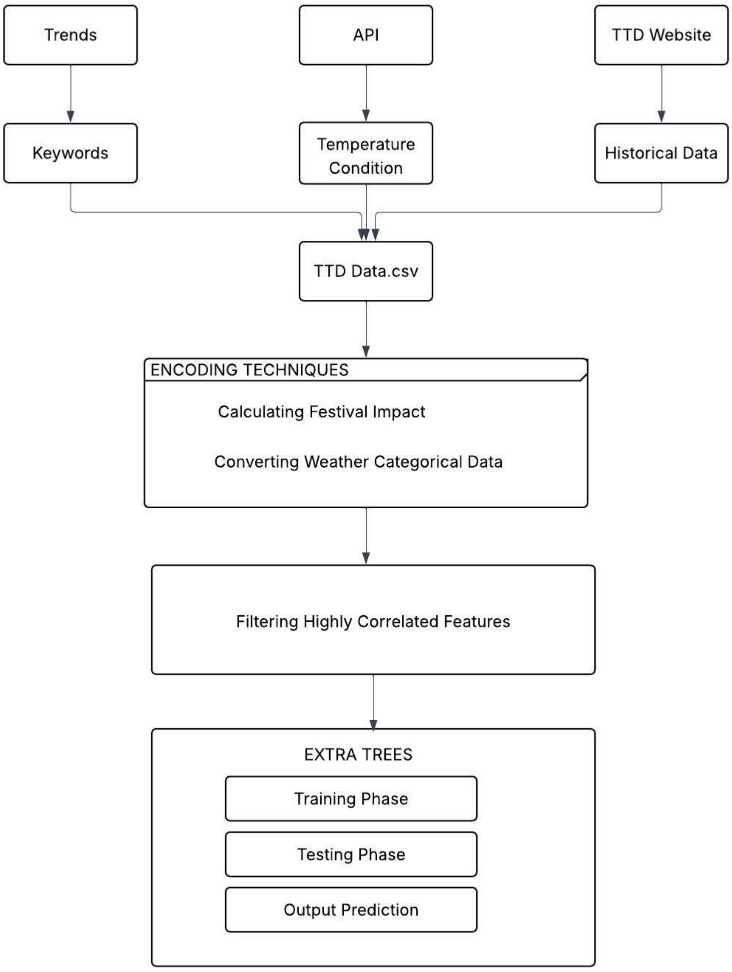


fig 1:The proposed approach to forecasting tourist arrival

Extra Trees Regressor Architecture:

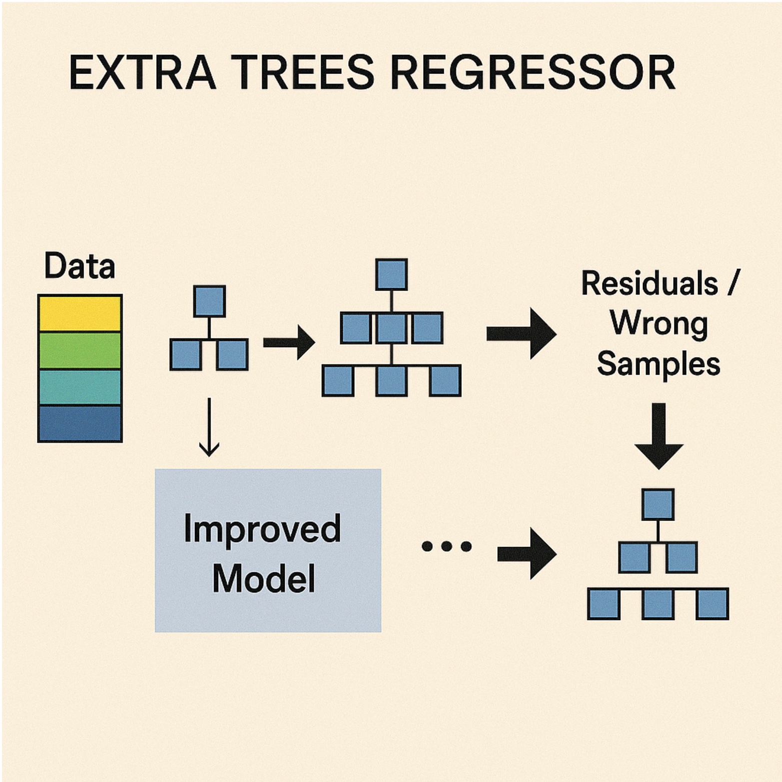


Fig 2: Extra Trees Regressor Model

The Extra Trees Regressor is an ensemble-based machine learning method that constructs multiple randomized decision trees and averages their predictions to enhance accuracy and prevent overfitting. It begins with a dataset comprising various features and the target variable. This input data is processed to generate numerous trees, with each tree built using random selections of both features and split thresholds.

Unlike traditional decision trees, the randomness introduced at both the feature and threshold levels ensures that each tree is diverse. These trees are trained independently and in parallel, without relying on previous models, making Extra Trees faster and more efficient than boosting algorithms. The ensemble of trees collectively contributes to a more generalized and stable model by averaging their predictions.

Though the concept of residuals or errors is more applicable to boosting models like LightGBM, it's sometimes referenced in visual diagrams for illustration. In Extra Trees, the final prediction is made by aggregating the outputs of all the trees, resulting in a robust and accurate model that is less prone to overfitting and effective in handling complex regression tasks.

Proposed Methodology for the Nearby Recommendation:

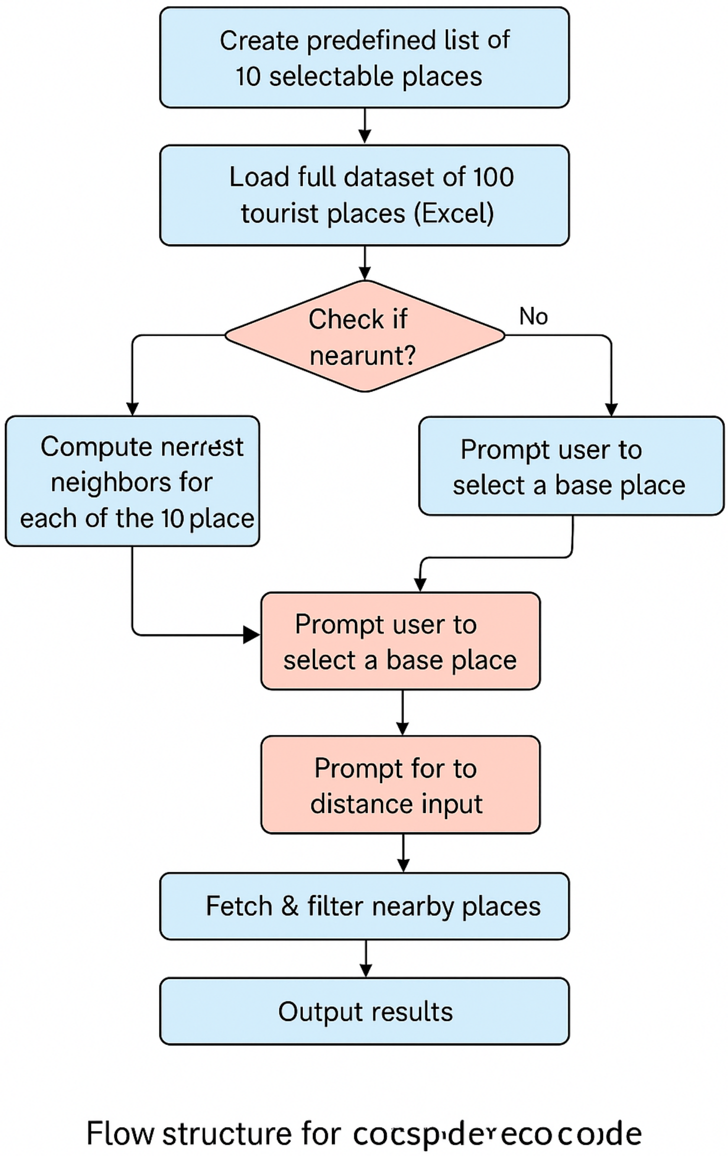


Fig 3: CKD Tree Flowchart

The flowchart outlines the structure of a nearby place recommendation system. It starts by creating a predefined list of 10 selectable tourist places and loading a complete dataset of 100 places from an Excel file. The system then checks if the "nearunt" (nearest neighbor) option is enabled. If it is, the nearest neighbors for each of the 10 places are computed. If not, the user is prompted to manually select a base location. After selecting a base place, the user is asked to input a distance value. The system then fetches and filters nearby places based on this input and outputs the final recommendations.

CKD Architecture:

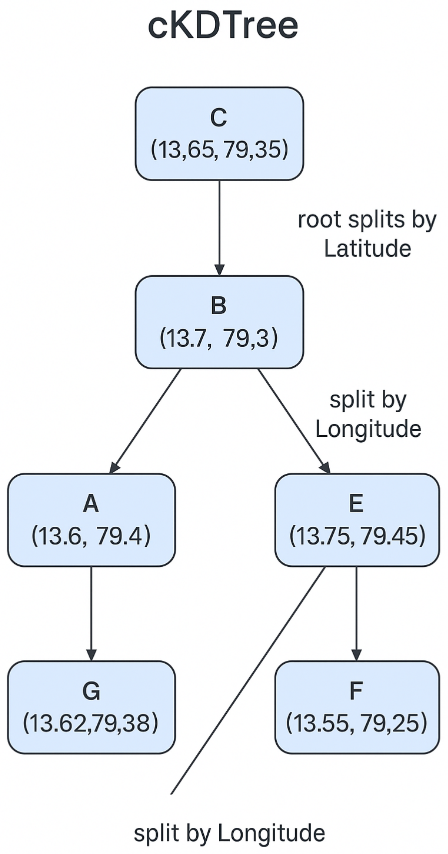


Fig 4: CKD model

The CKDTree is a binary tree structure used to efficiently organize and search geographic data points, like locations defined by latitude and longitude. It works by recursively splitting the dataset along different dimensions (e.g., latitude, longitude) at each level of the tree. The root level starts by dividing the data using one coordinate (like latitude), and the next level splits by the other coordinate (like longitude), and so on. This alternating split continues as the tree grows deeper.

This structure allows the system to quickly narrow down the search space when looking for nearby places. Instead of checking every point, the tree helps eliminate large sections of the data that are too far away, making nearest-neighbor searches much faster and more efficient.

Implementation for Forecasting

4.1 Data Collection

Data collection forms a critical foundation for this project, requiring considerable effort to gather and integrate information from multiple reliable sources into a unified, labeled dataset. To ensure the model is well-trained on historical patterns, data spanning the years 2022 to 2024 was utilized.

Pilgrim footfall statistics were sourced from the official Tirumala Tirupati Devasthanams (TTD) news portal (https://news.tirumala.org/), which provides daily updates on visitor counts. Complementary weather data was obtained from Weather Underground (https://www.wunderground.com/), offering daily atmospheric information.

Additionally, relevant keywords associated with Tirumala and its tourism ecosystem were identified, and corresponding Google Trends data was collected. This search trend data reflects public interest over time, thereby serving as a valuable indicator for visitor forecasting.

Since all datasets were organized on a daily basis, they were merged into a single, date-aligned dataset to ensure temporal consistency.

Data extraction from these online sources was automated using Selenium WebDriver in Python, enabling efficient and repeatable collection.

	Unnamed: 0	Date	Piligrims	Tonsures	Hundi	Tirupati	Tirumala	VIP darshan	tirumala darshan	srinivasam complex	tirupati balaji.ap.gov.in	temperature	weekday	conditions	Special_day
0	0	01-01-2022	36560	14084.0	2.15	43	23	16	9	0	0	71.60	1	Light drizzle	New Years Day
1	1	02-01-2022	38894	12270.0	3.93	41	25	14	9	0	0	74.84	1	Overcast	Amavasya, Sravanam, Tirupati Sri G.T Adhyayano...
2	2	03-01-2022	31776	16046.0	2.69	40	19	0	5	0	7	73.22	0	Moderate drizzle	Maha Sivaratri, Tirupati Sri KT Nandi Vahanam
3	3	04-01-2022	31523	14692.0	2.45	45	17	22	16	0	0	73.76	0	Overcast	Amavasya
4	4	05-01-2022	32044	17558.0	2.61	32	14	21	10	0	7	73.94	0	Overcast	May Day
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
1091	1091	27-12-2024	66715	24503.0	4.06	79	55	0	39	0	0	70.88	0	Moderate rain	normal
1092	1092	28-12-2024	78414	26100.0	3.45	76	50	0	28	0	0	76.46	1	Overcast	normal

Figure 5: Unpreprocessed Historical Data

This unprocessed dataset that I have collected captures daily pilgrim activity in **Tirupati** from **2022 to 2024**, featuring a rich blend of **numerical**, **categorical**, and **temporal data**. Key attributes include **pilgrim count**, **tonsures**, **hundi collections**, **entry counts at Tirupati and Tirumala**, and **VIP darshan** stats—offering insights into **crowd behavior** and **temple operations**. The inclusion of **weather conditions**, **temperature**, and **special religious days** adds valuable context for understanding **external influences**. With detailed **time-series information** and **event markers**, the dataset is well-suited for **forecasting**, **trend analysis**, and **resource planning** in religious tourism.

4.2 Feature Engineering (Data Preprocessing)

Similar to any **machine learning** project, a significant portion of the timeline is dedicated to **data preprocessing**. In our case, the dataset involves **real-time data** with various **unstructured formats**, making the task of transforming it into a clean and structured form particularly **labor-intensive**.

To handle categorical data, we applied **Target Encoding with Exponential Smoothing** to the **"Special Day"** column, generating a new feature called `festival_impact`. This technique accounts for the average pilgrim count per festival while incorporating the overall mean to ensure stability, especially for less frequent events.

Additionally, we performed **Label Encoding** on the **"conditions"** column, converting weather descriptions into numeric values using the following mapping:

```
{‘‘Clear sky’’: 0, ‘‘Dense drizzle’’: 1, ‘‘Heavy rain’’: 2, ‘‘Light
drizzle’’: 3, ‘‘Light rain’’: 4, ‘‘Mainly clear’’: 5, ‘‘Moderate
```

```
drizzle'': 6, ``Moderate rain'': 7, ``Overcast'': 8, ``Partly
cloudy'': 9}
```

These steps ensured that all **categorical values** were transformed into **numerical representations**, while existing **numerical features** were retained for modeling.

Overall, these preprocessing techniques were essential to prepare the dataset for **analysis** and **model training**, significantly contributing to the **accuracy** and **reliability** of the applied **machine learning algorithms**.

ndi	Tirupati	Tirumala	VIP darshan	tirumala darshan	srinivasam complex	tirupati balaji.ap.gov.in	temperature	Timestamp	weekday	festival_impact	holidays	Piligrims_7Day_Avg	condition_encoded
.15	43	23	16	9	0	0	71.60	1.640975e+09	1	42879.462226	1	36560.000000	3
.93	41	25	14	9	0	0	74.84	1.641062e+09	1	63674.217782	1	37727.000000	8
.69	40	19	0	5	0	7	73.22	1.641148e+09	0	68156.491023	0	35743.333333	6
.45	45	17	22	16	0	0	73.76	1.641235e+09	0	68156.491023	0	34688.250000	8
.61	32	14	21	10	0	7	73.94	1.641321e+09	0	68156.491023	0	34159.400000	8
...	...	...	...	...	...	...	...	...	...	...	...	...	...
.06	79	55	0	39	0	0	70.88	1.735238e+09	0	68156.491023	1	68876.571429	7
.45	76	50	0	28	0	0	76.46	1.735324e+09	1	68156.491023	1	69734.142857	8
.80	64	42	0	23	0	0	76.28	1.735411e+09	1	68156.491023	1	70832.714286	8
.10	68	42	0	26	0	0	74.66	1.735497e+09	0	68269.862226	1	71210.142857	3
.80	68	35	0	15	0	0	76.10	1.735583e+09	0	68156.491023	1	70536.714286	8

*****Figure 6:**processed Historical Data

After acquiring the raw dataset, comprehensive preprocessing was carried out to convert it into a clean and structured format. The processed dataset now maintains **consistent value ranges across all relevant features**, ensuring uniformity and eliminating irregularities present in the original unprocessed data.

Through meticulous encoding, normalization, and feature engineering—such as **target encoding** for festivals (resulting in the feature `festival_impact`) and **label encoding** for weather conditions—the dataset has been transformed into a format that is **well-aligned for modeling and analytical tasks**.

This cleaned version not only enhances **interpretability** but also improves the **efficiency and accuracy** of downstream **machine learning workflows**.

4.3 Selection Feature

In the context of the labelled dataset, there may exist redundant or unwanted features that can adversely affect the model's performance and computational efficiency. Feature selection is the process of selecting a subset of relevant features from a larger set of available features to improve the performance and efficiency of a machine learning model. The goal is to identify and retain the most informative features while discarding redundant or irrelevant ones.

To obtain relevant variables, we consider only those features that have a correlation with the output (i.e., pilgrims) greater than 0.2 or less than -0.2. Only these selected features are then provided to the algorithm for model construction.

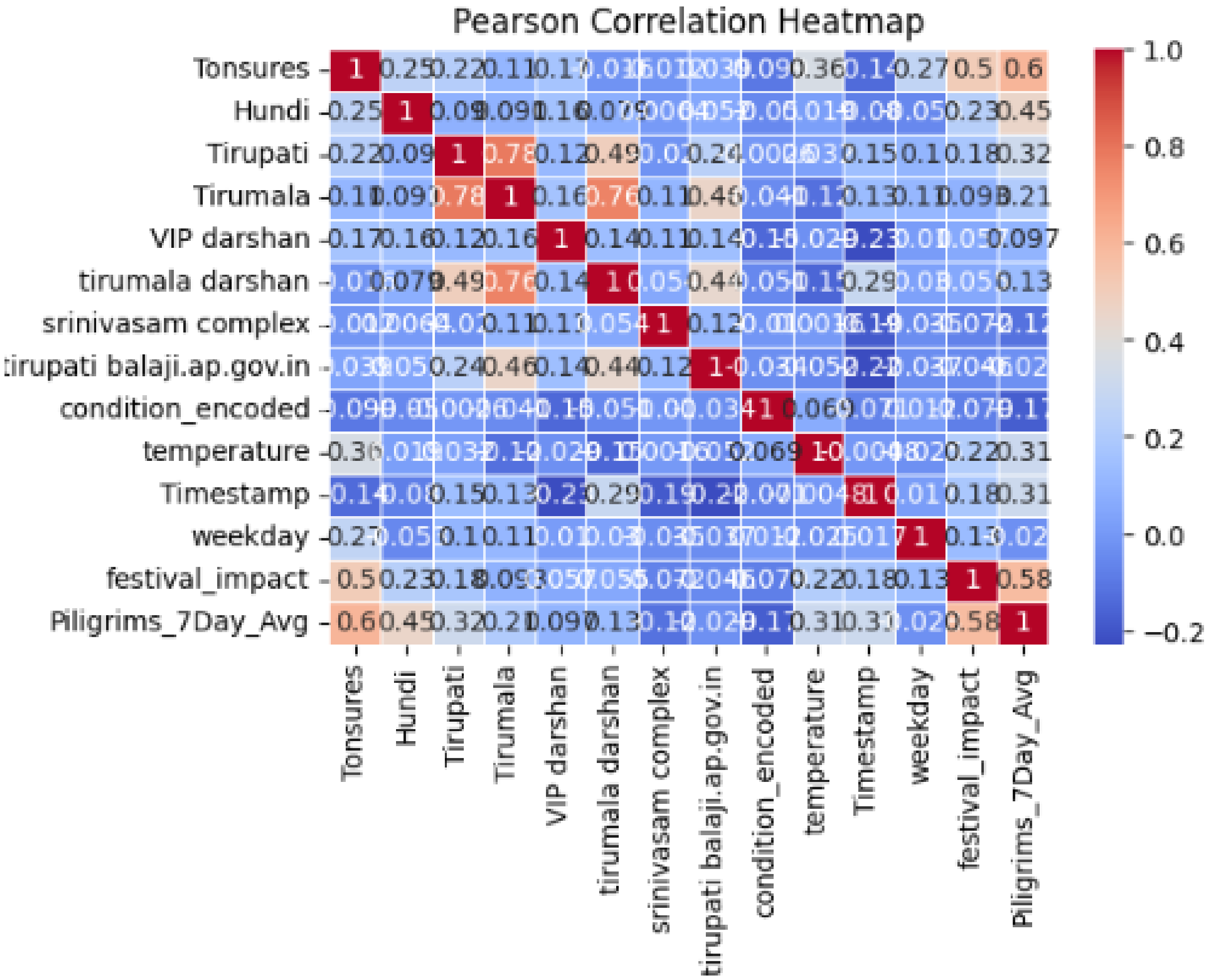


*__Figure 7:__Pearson Correlation Heatmap

```
Targe-independent features:   coefficient correlation value
------------------------------------------------------------
Piligrims-Tonsures: 0.6917
Piligrims-Hundi: 0.3691
Piligrims-Tirupati: 0.2906
Piligrims-Tirumala: 0.2033
Piligrims-temperature: 0.2417
Piligrims-Timestamp: 0.2323
Piligrims-weekday: 0.3941
Piligrims-festival_impact: 0.6320
Piligrims-Piligrims_7Day_Avg: 0.7620
```

*__Figure 8:__coefficient correlation value

4.4 Training Phase

The training phase of the **Extra Trees Regressor (ETR)** involves building an **ensemble** of fully grown, **randomized decision trees**. ETR introduces **randomness** by selecting random subsets of **features** and random **split points**, which helps reduce **variance** and enhances **generalization**.

Each tree is trained **independently** on the dataset, and the final prediction is computed by **averaging** the outputs of all trees. This approach makes ETR **robust** and effective for handling **high-dimensional** or **noisy** regression data.

Feature Importance Analysis

After training the Extra Trees Regressor model, we evaluated the contribution of each feature to the prediction of pilgrim counts using the model's built-in feature importance attribute. Feature importance provides insight into how valuable each feature is in predicting the target variable, based on how much they reduce impurity across the ensemble of trees.

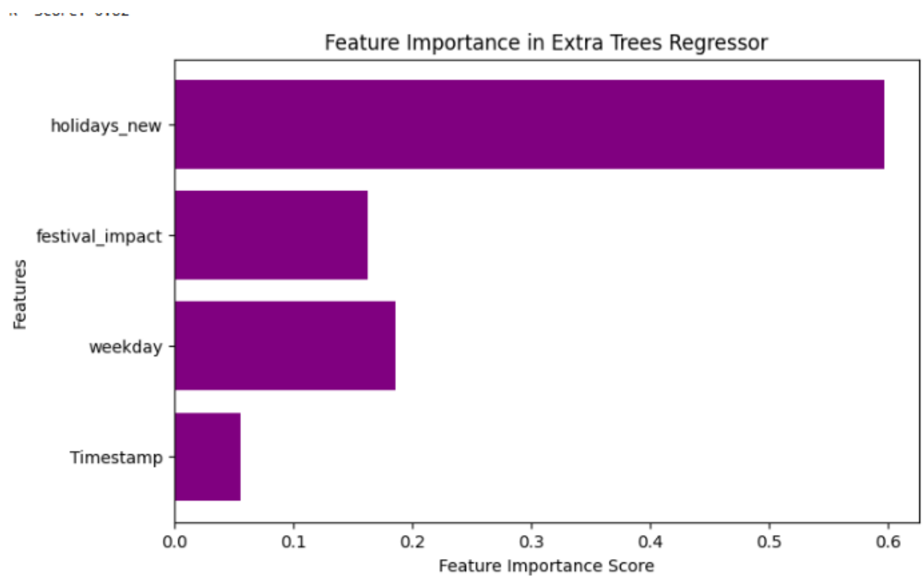


***Figure 9:**Feature Importance

The feature importance analysis shows that `holidays_new` is the most impactful feature, strongly influencing pilgrim counts. `festival_impact` and `weekday` also contribute moderately, indicating the importance of special days and weekly patterns. In contrast, `Timestamp` has minimal impact, suggesting that raw time data is less useful without further transformation. Overall, the model relies most on event-based features to predict pilgrim trends.

4.5 Validation Phase

The validation phase for the **Extra Trees Regressor (ETR)** involves evaluating the performance of the trained model on a separate **validation dataset**. This phase is crucial for assessing the model's ability to **generalize** to unseen data and for identifying potential issues such as **overfitting**.

The ETR model generates **predicted values** for the input features in the validation set, which are then compared to the corresponding **true values** of the target variable. This comparison enables the computation of various **performance metrics** such as Mean Squared Error (MSE), Mean Absolute Error (MAE), or R-squared (R^2), providing a quantitative measure of the model's **predictive accuracy**.

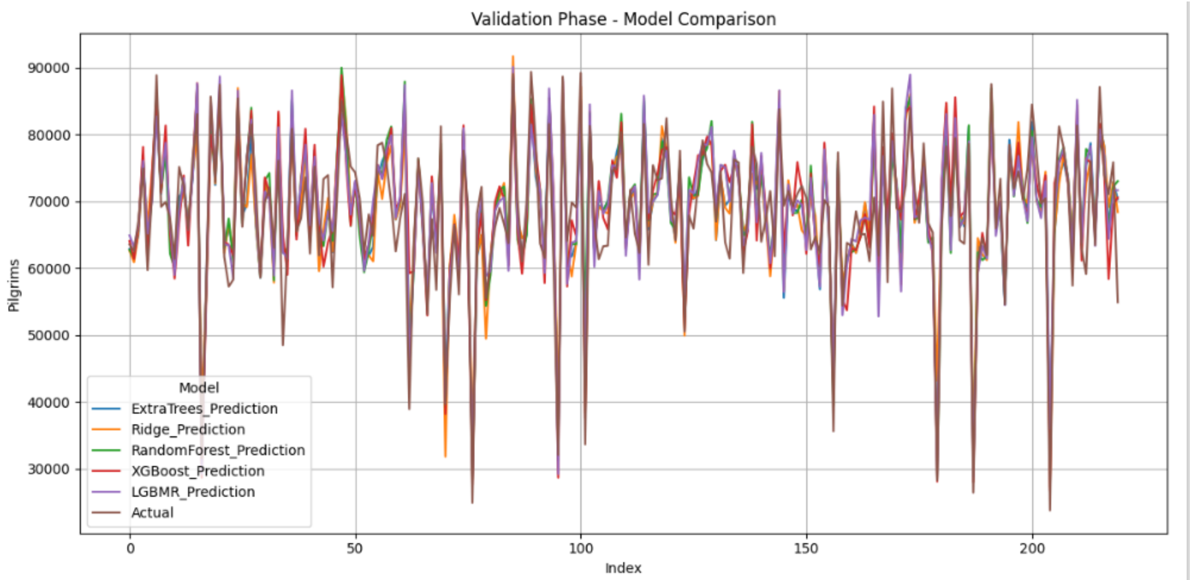


***Figure 9:**Validation Phase

Implementation for the Nearby Tourists Places

Dataset collection

To enhance the system's utility beyond footfall prediction, a complementary dataset was developed to provide detailed information about tourist attractions located in and around Tirupati and Tirumala. This dataset serves as the foundation for a proximity-based recommendation system, which suggests nearby places of interest based on the user's selected base location and preferred travel radius.

Source: The dataset was manually compiled and structured into an Excel file. It contains a total of **100 unique tourist destinations**, each described with relevant geographical and categorical attributes.

	Place Name	God/Goddess/Attraction	Category	Latitude	Longitude
0	Sri Venkateswara Temple, Tirumala	Lord Venkateswara	Devotional	13.6833	79.3474
1	Sri Padmavathi Ammavari Temple, Tiruchanur	Goddess Padmavathi	Devotional	13.6078	79.4501
2	Sri Govindaraja Swamy Temple, Tirupati	Lord Govindaraja	Devotional	13.6833	79.3500
3	Sri Kapileswara Swamy Temple, Tirupati	Lord Shiva	Devotional	13.6400	79.4000
4	ISKCON Temple, Tirupati	Lord Krishna & Radha	Devotional	13.6311	79.4192
5	Sri Kalahasteeswara Temple, Srikalahasti	Lord Shiva (Vayu Lingam)	Devotional	13.7490	79.7037
6	Kanipakam Vinayaka Temple, Kanipakam	Lord Ganesha	Devotional	13.3350	79.0670
7	Srinivasa Mangapuram Temple, Mangapuram	Lord Venkateswara	Devotional	13.6167	79.3167
8	Chandragiri Fort, Chandragiri	Historical Fort	History	13.5833	79.3333
9	Silathoranam, Tirumala	Natural Rock Formation	Nature	13.6833	79.3500
10	Talakona Waterfalls, Chittoor District	Natural Waterfall	Nature	13.8500	79.1500
11	Sri Venkateswara National Park, Chittoor	Wildlife Reserve	Nature	13.7167	79.3500
12	Sri Venkateswara Zoological Park, Tirupati	Zoo & Wildlife Sanctuary	Nature	13.6288	79.4192
13	Chakra Theertham, Tirumala	Holy Water Body	Holy Site	13.6833	79.3500
14	Akasa Ganga Theertham, Tirumala	Holy Waterfall	Holy Site	13.6833	79.3500
15	Papavinasanam Theertham, Tirumala	Holy Waterfall	Holy Site	13.6833	79.3500
16	Kapila Theertham Waterfall, Tirupati	Natural Waterfall	Nature	13.6400	79.4000
17	Deer Park, Tirumala	Wildlife Park	Nature	13.6833	79.3500
18	Japali Theertham, Tirumala	Lord Hanuman	Holy Site	13.6833	79.3500
19	Gudimallam Temple, Gudimallam	Lord Shiva	Devotional	13.5167	79.4500
20	Alipiri Mettu (Trek Route), Tirupati	Footpath to Tirumala	Pilgrim Route	13.6300	79.4200

Fig 10:Nearby Places

Training Phase

In this system, a traditional machine learning model is not employed. Instead, the core logic is built upon a geospatial proximity-based model utilizing the `cKDTree` algorithm from the `SciPy` library.

During the training-like phase, the geographical coordinates (latitude and longitude) of 100 tourist locations are used to construct a `cKDTree`, enabling efficient nearest-neighbor searches. For each predefined tourist destination, the system identifies surrounding places within a user-defined radius and computes the precise distances using the `geopy` library, which accounts for the curvature of the Earth to ensure accuracy.

The resulting neighborhood mappings and distances are cached and stored in a structured format for rapid access. This preprocessing phase greatly enhances performance during the inference stage, enabling real-time, location-based recommendations without repeated distance calculations.

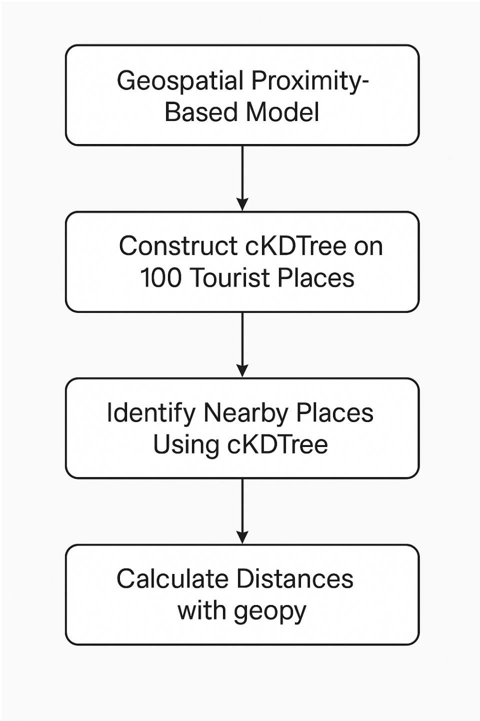


Fig 11: Flow Diagram

Including Google maps

Applications: This dataset enables dynamic recommendations using spatial filtering techniques. It was integrated into the application using geospatial libraries such as `Folium` and `MarkerCluster`, allowing users to explore attractions interactively on a digital map based on proximity and interest.

Development of Web Application

Web Interface

The web interface serves as the homepage for my Tirumala Tirupati Devasthanam application, providing a brief introduction about the temple's spiritual and cultural significance. It features two main options — "Find Pilgrims" to predict the number of visitors, and "Find Nearby Places" to explore surrounding attractions, offering a user-friendly entry point for pilgrims and tourists.

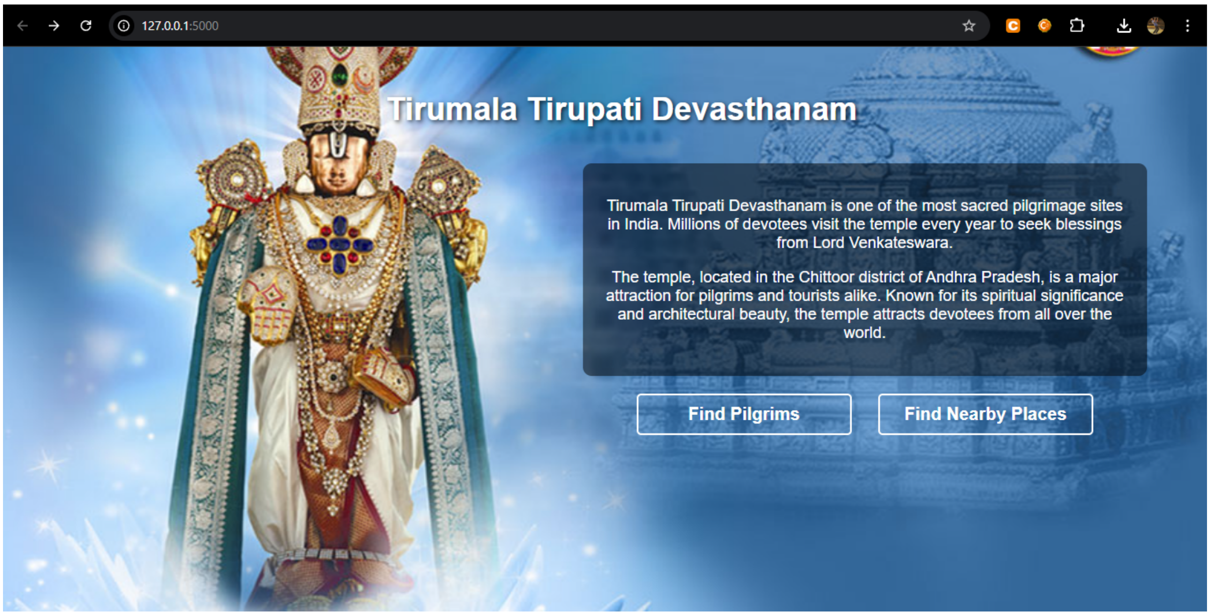


Fig 12: Web Interface

Output Prediction

The pilgrimage prediction system is designed to estimate the number of pilgrims expected on a specific day based on user inputs such as date, weekend or holiday status, and festival name. It leverages the **Extra Trees Regression** model, an ensemble learning technique that constructs multiple decision trees and averages their outputs for improved accuracy and consistency. This helps temple authorities with better planning and crowd control.

The user-friendly web interface allows individuals to easily provide relevant input details and, with a simple click of the "Predict" button, instantly receive the estimated number of pilgrims. The model processes the data and displays the prediction on the same page, offering a seamless and efficient experience for users and stakeholders involved in managing temple logistics.

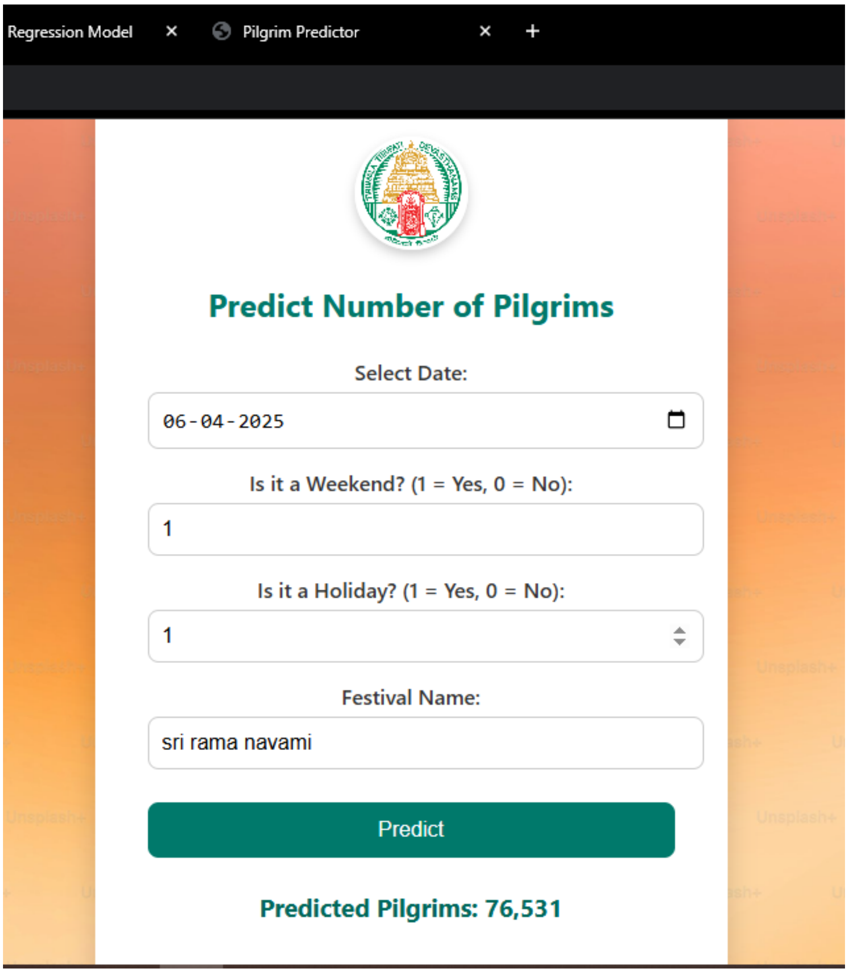
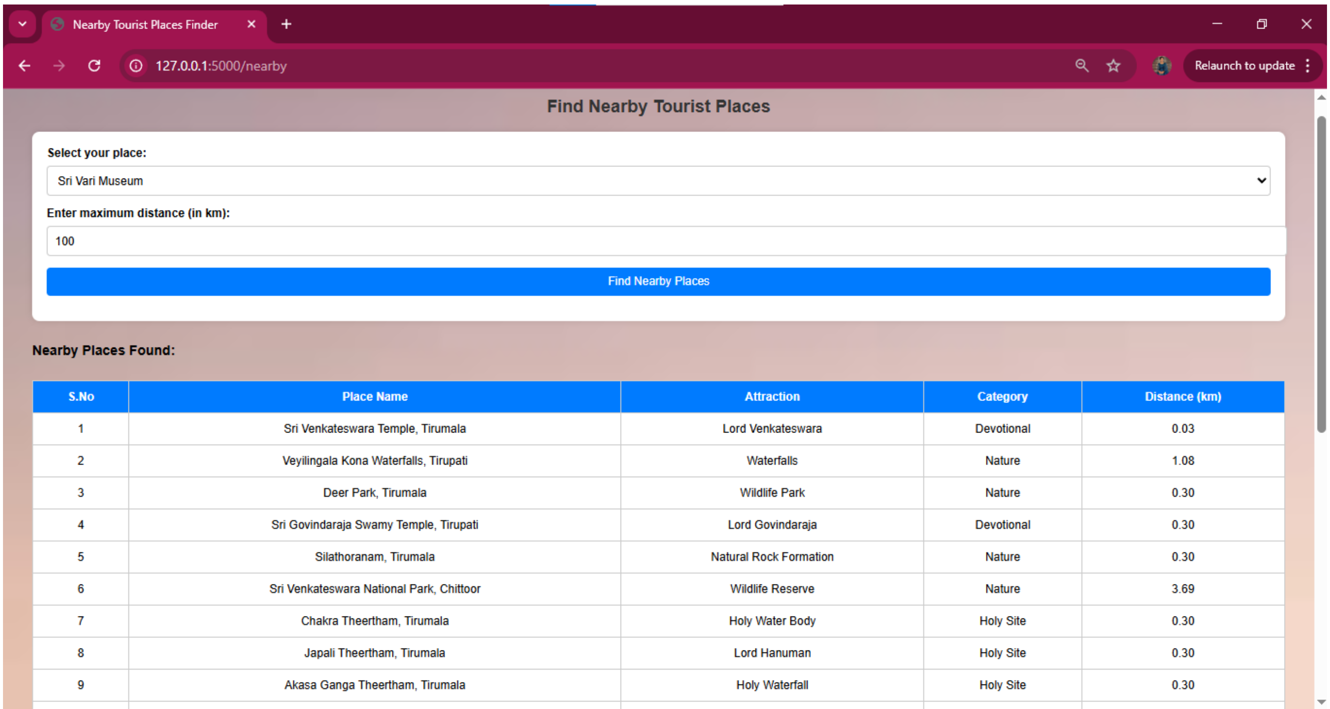


Fig 13: Output

Nearby recommending

Application also enables users to find tourist places near a selected location. Users can choose a place from a dropdown menu and specify a maximum distance in kilometers. Once the user clicks the "Find Nearby Places" button, the system fetches and displays a list of nearby attractions within the given radius.

The resulting table includes essential details such as rank, place name, key attraction, category (like devotional, nature, or wildlife), and distance from the selected location. This feature helps users plan their visit more effectively by identifying multiple points of interest close to their destination.



S.No	Place Name	Attraction	Category	Distance (km)
1	Sri Venkateswara Temple, Tirumala	Lord Venkateswara	Devotional	0.03
2	Veyilingala Kona Waterfalls, Tirupati	Waterfalls	Nature	1.08
3	Deer Park, Tirumala	Wildlife Park	Nature	0.30
4	Sri Govindaraja Swamy Temple, Tirupati	Lord Govindaraja	Devotional	0.30
5	Silathoranam, Tirumala	Natural Rock Formation	Nature	0.30
6	Sri Venkateswara National Park, Chittoor	Wildlife Reserve	Nature	3.69
7	Chakra Theertham, Tirumala	Holy Water Body	Holy Site	0.30
8	Japali Theertham, Tirumala	Lord Hanuman	Holy Site	0.30
9	Akasa Ganga Theertham, Tirumala	Holy Waterfall	Holy Site	0.30

Fig 14 :Nearby Places

Google Maps Output

Once the user selects a base location and specifies a distance, the map dynamically marks the base and nearby places, providing a spatial understanding of their proximity. The map enhances user experience by enabling easier route planning and orientation, especially for tourists who want to explore multiple attractions within a short distance. This feature significantly improves the utility of the recommendation system by integrating geographic context into the decision-making process.

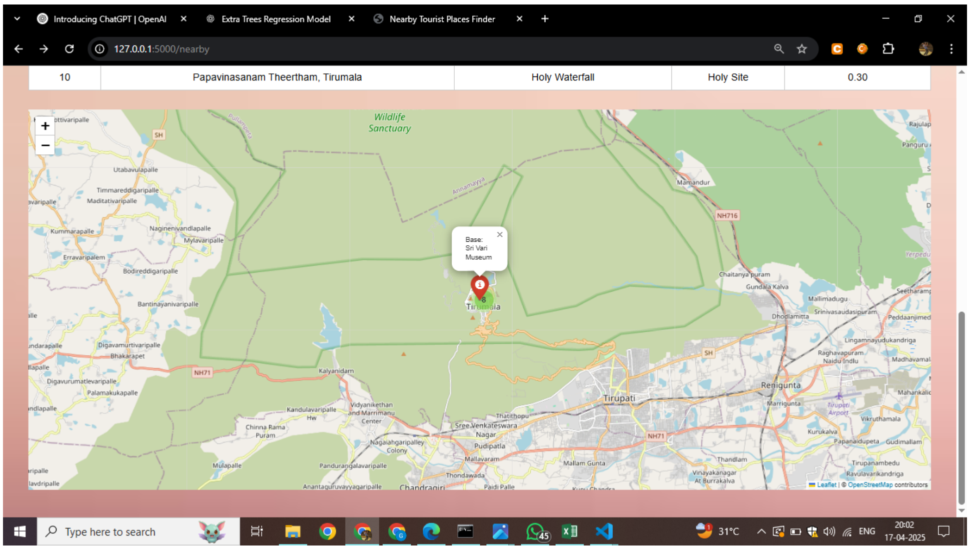


Fig 14 : Google Maps

Dataset Description

Dataset for Forecasting Tourist Visits

The dataset contains various features that help predict the number of pilgrims visiting Tirupati and Tirumala on a given day. It includes weather, temporal, and event-based variables such as holidays and festivals. This dataset is used to train a machine learning model (e.g., Extra Trees Regressor) for estimating pilgrim flow and assisting authorities in effective planning and management.

date: The actual calendar date corresponding to each record in the dataset.The data is from 2022-2024

hundis: Amount of money collected in temple hundis (donation boxes), indicating the donation volume.

tonsures: Number of head tonsuring rituals performed on that day, often a major religious act by pilgrims.

Tirupati: Represents the search count using google trends for the keyword Tirupati on a specific day.

Tirumala: Represents the search count using google trends for the keyword Tirumala on a specific day.

VIP darshan: Indicates the count of pilgrims who availed VIP darshan services on that day.

Tirumala darshan: Refers to the general darshan (regular entry) count in Tirumala.

Srinivasam complex: Number of pilgrims staying or utilizing facilities at the Srinivasam Complex.

Tirupati balaji.ap.gov.in: Number of bookings or interactions from the official online portal.

Temperature: Temperature (in F) on that particular day, which may influence crowd behavior.

ı**Timestamp:** A Unix timestamp representing the exact date and time of the data entry.

Weekday: Indicates whether the day is a weekened (1) or not (0), used to identify patterns.

Festival impact: A numerical score estimating the influence of festivals on the expected footfall.

Holidays: A binary value showing whether the day is a public holiday (1) or not (0).

Pilgrims_7Day_Avg: The 7-day moving average of pilgrims, used to smooth and analyze trends.

Condition_encoded: Encoded values representing categorical variables such as weather conditions,

ndi	Tirupati	Tirumala	VIP darshan	tirumala darshan	srinivasam complex	tirupati balaji.ap.gov.in	temperature	Timestamp	weekday	festival_impact	holidays	Piligrims_7Day_Avg	condition_encoded
2.15	43	23	16	9	0	0	71.60	1.640975e+09	1	42879.462226	1	36560.000000	3
3.93	41	25	14	9	0	0	74.84	1.641062e+09	1	63674.217782	1	37727.000000	8
2.69	40	19	0	5	0	7	73.22	1.641148e+09	0	68156.491023	0	35743.333333	6
2.45	45	17	22	16	0	0	73.76	1.641235e+09	0	68156.491023	0	34688.250000	8
2.61	32	14	21	10	0	7	73.94	1.641321e+09	0	68156.491023	0	34159.400000	8
...	...	...	...	...	...	...	...	...	...	...	...	...	...
4.06	79	55	0	39	0	0	70.88	1.735238e+09	0	68156.491023	1	68876.571429	7
3.45	76	50	0	28	0	0	76.46	1.735324e+09	1	68156.491023	1	69734.142857	8
3.80	64	42	0	23	0	0	76.28	1.735411e+09	1	68156.491023	1	70832.714286	8
4.10	68	42	0	26	0	0	74.66	1.735497e+09	0	68269.862226	1	71210.142857	3
3.80	68	35	0	15	0	0	76.10	1.735583e+09	0	68156.491023	1	70536.714286	8

Fig 15: Entire Dataset

Dataset for Nearby Attractions

This dataset contains a collection of nearby places around Tirumala and Tirupati, including temples, waterfalls, historical sites, and nature reserves. It is useful for tourism, pilgrimage planning, and mapping applications.

Place Name – Name of the temple, site, or attraction near Tirupati/Tirumala.

God/Goddess/Attraction – The deity or main point of interest at each location.

Category – Type of location (e.g., Devotional, Nature, History, Holy Site).

Latitude – Geographic latitude used to map the place.

Longitude – Geographic longitude used to map the place.

	Place Name	God/Goddess/Attraction	Category	Latitude	Longitude
0	Sri Venkateswara Temple, Tirumala	Lord Venkateswara	Devotional	13.6833	79.3474
1	Sri Padmavathi Ammavari Temple, Tiruchanur	Goddess Padmavathi	Devotional	13.6078	79.4501
2	Sri Govindaraja Swamy Temple, Tirupati	Lord Govindaraja	Devotional	13.6833	79.3500
3	Sri Kapileswara Swamy Temple, Tirupati	Lord Shiva	Devotional	13.6400	79.4000
4	ISKCON Temple, Tirupati	Lord Krishna & Radha	Devotional	13.6311	79.4192
5	Sri Kalahasteeswara Temple, Srikalahasti	Lord Shiva (Vayu Lingam)	Devotional	13.7490	79.7037
6	Kanipakam Vinayaka Temple, Kanipakam	Lord Ganesha	Devotional	13.3350	79.0670
7	Srinivasa Mangapuram Temple, Mangapuram	Lord Venkateswara	Devotional	13.6167	79.3167
8	Chandragiri Fort, Chandragiri	Historical Fort	History	13.5833	79.3333
9	Silathoranam, Tirumala	Natural Rock Formation	Nature	13.6833	79.3500
10	Talakona Waterfalls, Chittoor District	Natural Waterfall	Nature	13.8500	79.1500
11	Sri Venkateswara National Park, Chittoor	Wildlife Reserve	Nature	13.7167	79.3500
12	Sri Venkateswara Zoological Park, Tirupati	Zoo & Wildlife Sanctuary	Nature	13.6288	79.4192
13	Chakra Theertham, Tirumala	Holy Water Body	Holy Site	13.6833	79.3500
14	Akasa Ganga Theertham, Tirumala	Holy Waterfall	Holy Site	13.6833	79.3500
15	Papavinasanam Theertham, Tirumala	Holy Waterfall	Holy Site	13.6833	79.3500
16	Kapila Theertham Waterfall, Tirupati	Natural Waterfall	Nature	13.6400	79.4000
17	Deer Park, Tirumala	Wildlife Park	Nature	13.6833	79.3500
18	Japali Theertham, Tirumala	Lord Hanuman	Holy Site	13.6833	79.3500
19	Gudimallam Temple, Gudimallam	Lord Shiva	Devotional	13.5167	79.4500
20	Alipiri Mettu (Trek Route), Tirupati	Footpath to Tirumala	Pilgrim Route	13.6300	79.4200

Fig 16 : Entire Dataset

Results

Forecasting Tourist Visits

Our tourist demand prediction model for Tirumala Tirupati Devasthanam (TTD). The dataset used for training and evaluation included features such as Google Trends data, historical data, day type, and weather conditions.

We employed five different machine learning algorithms — Extra Trees Regressor, Random Forest, XGBoost, and LightGBM — to predict tourist demand for TTD. Each model was trained and evaluated using the collected dataset.

To assess the performance of the models, we used standard evaluation metrics including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) value. These metrics help quantify the accuracy and predictive strength of each model.

The results of our analysis revealed strong predictive performance across all five models. Among them, Extra trees regression consistently delivered the lowest MAE and RMSE, indicating the highest accuracy. Additionally, it achieved the highest R^2 score, demonstrating a strong correlation between predicted and actual tourist demand.

Overall, the findings from our tourist demand prediction model show promising potential for accurately forecasting future tourist flow to Tirumala Tirupati Devasthanam (TTD). While the current models perform well, further refinement and real-time data integration could further enhance their accuracy and robustness.

These insights can support TTD in making data-driven decisions regarding resource allocation, crowd management, and visitor service planning. By effectively forecasting tourist demand, TTD can optimize its operations and improve the overall experience for its pilgrims and visitors.

Model Comparison Matrix:

	MAE	RMSE	R^2
Extra Trees	3331.5079	4263.1691	0.8630
Random Forest	3374.8355	4393.8314	0.8545
LightGBM	3464.6735	4468.8260	0.8494
XGBoost (Optimized)	3568.4122	4488.6556	0.8481

Fig 17: Comparision Table

Nearby Recommendations

To enhance the pilgrim experience and assist visitors in exploring the surrounding areas of Tirumala and Tirupati, we implemented a nearby place recommendation system using the CKD (Compressed K-Dimensional) Tree model. By leveraging the geographical coordinates (latitude and longitude) of various devotional, natural, and historical sites, the model efficiently identifies and suggests the nearest relevant attractions based on a visitor's current or selected location. This spatial query approach significantly improves recommendation accuracy and speed, especially in scenarios with large datasets. The CKD Tree model proved to be an effective solution for proximity-based recommendations, supporting TTD in offering pilgrims contextual guidance and enriching their overall visit with well-informed exploration options around the region.

Requirements

Software Requirements

- Jupyter Notebook
- Operating System: Windows
- Technology: Python 3
- Packages: `pandas`, `numpy`, `scikit-learn`, `xgboost`, `lightgbm`, `matplotlib`, `seaborn`, `tensorflow`, `DecisionTree`, `ExtraTreesRegressor`

Hardware Requirements

- Processor: AMD Ryzen 3 3250U with Radeon Graphics @ 2.60 GHz
- Installed RAM: 8.00 GB (5.94 GB usable)
- System Type: 64-bit Operating System, x64-based Processor
- Storage: Minimum 5 GB free disk space

Future Scope

Tourist Demand Prediction Enhancements

- **Integration of Additional Social Networking Data:** Enrich the model using social media platforms such as *Instagram*, *Facebook*, and *Twitter*. By performing sentiment analysis, extracting user-generated content, and tracking engagement metrics, the model can better understand tourist behavior and trends.

- **Prediction for Multiple Destinations:** Extend the model to support multi-location forecasting beyond Tirumala Tirupati Devasthanam (TTD). While current limitations exist due to dataset availability, future expansions could enable broader regional tourism analysis.

- **Collaboration with Tourism Stakeholders:** Partner with tourism boards, travel agencies, and industry professionals to collect richer datasets and refine prediction goals. Stakeholder insights and proprietary data could significantly boost model accuracy and its practical deployment in real-world planning.

Nearby Recommendations Using Google Maps

- **Dynamic Personalization Based on User Behavior:** Future iterations can personalize nearby recommendations by analyzing user interests, time of visit, walking/driving preferences, and past behavior to tailor results.

- **Real-time Data Integration:** Integrating real-time data from the Google Maps API, such as live traffic conditions, opening hours, or crowd density, could refine the recommendations and offer more context-aware suggestions.

- **Multi-Criteria Ranking of Places:** Enhance the recommendation engine to rank nearby places not just by distance, but also by user reviews, popularity, category (e.g., temples, food spots), and historical significance.

- **Voice-Enabled and Mobile Optimization:** Implement voice-activated search and develop a mobile-friendly interface or app version of the recommendation system, making it easier for on-the-go tourists to access relevant suggestions instantly.

- **Offline Functionality for Remote Areas:** To support tourists with limited internet access, incorporate offline map caching and basic place recommendations based on preloaded data within a mobile app.

Conclusion

This project aimed to build a comprehensive system for predicting tourist demand and recommending nearby places for visitors to **Tirumala Tirupati Devasthanam (TTD)**. Using machine learning techniques such as *LightGBM, XGBoost, Extra Trees Regressor, Random Forest, and Ridge Regression*, we successfully developed a prediction model that achieved high accuracy in forecasting pilgrim inflow based on features like historical data, weather, festivals, and online trends. Among the models, **Extra trees Regression** outperformed others in terms of evaluation metrics, demonstrating strong reliability and potential for practical deployment.

In addition, we developed a **nearby place recommendation system** using a *CKD Tree-based spatial search*, integrated with **Google Maps**, to assist pilgrims in exploring relevant devotional and cultural landmarks around Tirumala and Tirupati. This feature adds value by enhancing visitor engagement and improving the overall tourism experience.

Together, these systems provide a powerful toolset for **TTD administrators, tourism planners, and pilgrims**, offering both predictive insights and real-time guidance. The integration of data-driven predictions with spatial recommendations has the potential to significantly improve *resource planning, crowd management, and service optimization* in religious tourism.

With further enhancements such as the inclusion of social media data, multi-destination predictions, and mobile accessibility, this project can evolve into a robust, scalable solution that benefits not only TTD but also the broader tourism ecosystem.

References

1. Haiyan Song, Stephen F. Witt,
   Tourism demand modelling and forecasting – A review of recent research, Tourism Management, Vol. 15, No. 5 (1995), pp. 409–421.
   https://www.researchgate.net/publication/$238087247_{T}ourism_{d}emand_{m}odelling_{a}nd_{f}orecasting_{-A}$

2. *The evolution of tourism and tourism research*, Tourism Recreation Research .
   https://www.researchgate.net/publication/$276428368_{T}he_{e}volution_{o}f_{t}ourism_{a}nd_{t}ourism_{r}esearch$