

TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
ADVANCED COLLEGE OF ENGINEERING AND MANAGEMENT
DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING
KALANKI, KATHMANDU



**A Minor Project Proposal Defense Report On
“MACHINE LEARNING–BASED CAFE LOCATION AND CAFE
TYPE SUITABILITY PREDICTION SYSTEM”**

[CT-654]

Submitted By:

Santosh Mahato Koiri / ACE079BCT059

Sijan Shrestha / ACE079BCT065

Upendra Khadka / ACE079BCT077

A Minor Project Proposal report submitted to the Department of Electronics and Computer Engineering in the partial fulfillment of the requirements for Degree of Bachelor of Engineering in Computer Engineering
Kathmandu, Nepal

5th February, 2026

Table of Contents

List of Tables	iii
List of Figures	iv
List of Abbreviations	v
CHAPTER 1	1
INTRODUCTION	1
1.1 Background.....	1
1.2 Motivation.....	1
1.3 Problem Statements.....	2
1.4 Project Objective.....	3
1.5 Significance of the Study.....	3
CHAPTER 2	4
LITERATURE REVIEW	4
2.1 Related Works.....	4
2.2 Summary of Literature Review.....	5
2.3 Research Gap	6
CHAPTER 3	7
REQUIREMENT ANALYSIS	7
3.1 Functional Requirements.....	7
3.2 Non-Functional Requirements.....	7
3.3 System Requirements	8
3.3.1 Hardware Requirements.....	8
3.3.2 Software Requirements	8
3.4 Feasibility Analysis	8
3.4.1 Technical Feasibility	8
3.4.2 Economic Feasibility	9
3.4.3 Operational Feasibility	9
CHAPTER 4	10
SYSTEM DESIGN AND ARCHITECTURE.....	10
4.1 System Architecture Design	10
4.2 Use Case Diagram.....	13
4.3 Data Flow Diagram (DFD).....	13

4.4 Flowchart for the Workflow of Project	14
CHAPTER 5	17
METHODOLOGY	17
5.1 Multi-Source Data Acquisition	17
5.2 Spatial Feature Engineering and Buffer Analysis.....	18
5.3 Multi-Criteria Suitability Analysis (MCSA)	19
5.4 Model Training	19
5.5 Evaluation Metrics	20
5.6 Software Development Model: Agile Model	20
CHAPTER 6	22
EXPECTED OUTPUT	22
6.1 Interactive Web Interface	22
6.2 Spatial Analysis Results	22
6.3 Machine Learning Based Recommendations	23
CHAPTER 7	24
TIME SCHEDULE.....	24
7.1 Project Timeline Overview.....	24
7.2 Gantt Chart.....	24
7.3 Detailed Task Breakdown	24
REFERENCES	26

List of Tables

Table 1: Literature Review	4
Table 2: Evaluation Metrics	20
Table 3: Gantt Chart	24

List of Figures

Figure 1: Design of System Architecture	10
Figure 2: User Interface Layer	11
Figure 3: Application Layer.....	11
Figure 4: Machine Learning Layer.....	12
Figure 5: Data Layer.....	13
Figure 6: Use Case Diagram for Overall Operation.....	14
Figure 7: DFD Level 0.....	15
Figure 8: DFD Level 1.....	15
Figure 9: Flowchart of the System	16
Figure 10: Methodology Overview	17
Figure 11: Agile Software Development Model	21
Figure 12: Expected Output Web Interface	22

List of Abbreviations

AHP	Analytic Hierarchy Process
API	Application Programming Interface
CSV	Comma-Separated Values
DFD	Data Flow Diagram
GB	Gigabyte
GCN	Graph Convolutional Networks
GIS	Geographic Information System
HTTP	Hypertext Transfer Protocol
JSON	JavaScript Object Notation
LBRS	Location-Based Recommender Systems
MCDM	Multi-Criteria Decision Making
MCSA	Multi-Criteria Suitability Analysis
ML	Machine Learning
NSO	National Statistics Office (Nepal)
OAuth	Open Authorization
OSM	OpenStreetMap
PROMETHEE	Preference Ranking Organization Method for Enrichment of Evaluations
RAM	Random Access Memory
UI/UX	User Interface / User Experience
WGS84	World Geodetic System 1984

CHAPTER 1

INTRODUCTION

1.1 Background

Choosing the physical location for a business is a decision of strategic importance that significantly influences the survival and performance of a company [1]. In the hospitality sector, particularly for cafes and restaurants, location is often considered the most critical determinant of success. Traditionally, these decisions were based on intuition or basic observation, weighing factors such as population density, local economic conditions, and accessibility [1]. However, with the rapid urbanization of cities like Kathmandu, the complexity of these variables has increased, transforming location selection into a multi-criteria decision problem that requires sophisticated analysis rather than simple estimation.

The emergence of Location-Based Recommender Systems (LBRS) has provided a technological solution to this challenge. LBRS are specialized context-aware systems that utilize geographical data to provide users with timely and relevant items or information [3]. By integrating Geographic Information Systems (GIS) with machine learning techniques, it is now possible to process vast amounts of spatial data such as competitor locations, road networks, and demographic data to generate actionable business insights [4]. While traditional recommender systems focus on suggesting products to users, a business-oriented LBRS focuses on recommending the optimal location or business type for a specific area.

This project proposes a web-based application designed to assist entrepreneurs in Kathmandu Metropolitan City. By leveraging spatial data from OpenStreetMap and Google Maps, combined with demographic data, the system analyzes the "business influence area" to evaluate suitability and predict the success of new cafe ventures using machine learning algorithms like Random Forest.

1.2 Motivation

The primary motivation for this project stems from the high failure rate of new establishments in the food and beverage industry. Statistics indicate a harsh reality where a significant percentage of restaurants fail within their first year, often due to suboptimal location choices and an inability to adapt to the local market saturation [2].

Chain restaurants may afford to close underperforming outlets, but for small-scale entrepreneurs in Kathmandu, a wrong location choice can lead to irreversible financial loss.

Furthermore, there is a lack of accessible, data-driven tools for local investors. Most existing systems are descriptive rather than prescriptive [2]. There is a growing need for "Prescriptive Analytics" in business location selection systems that not only analyze the current state but also recommend the best course of action [2]. This project is motivated by the desire to bridge this gap, providing a user-friendly platform that democratizes access to complex spatial analysis and machine learning predictions, ultimately fostering a more sustainable business ecosystem in Kathmandu.

1.3 Problem Statements

Despite the availability of map services like Google Maps, entrepreneurs in Kathmandu face significant challenges in assessing the viability of a potential cafe location. The core problem lies in the difficulty of manually aggregating and analyzing dispersed spatial data to make an informed decision. Specifically, the problems addressed by this project include:

a) Information Asymmetry:

Entrepreneurs often lack a consolidated view of critical parameters such as competitor density, population demographics, and road accessibility within a specific radius of a proposed site.

b) Subjectivity in Decision Making:

Without a quantitative system, decisions are often driven by intuition, leading to businesses opening in oversaturated areas or missing opportunities in underserved locations [1].

c) Lack of Predictive Insight:

While a user might know where existing cafes are, they often lack the analytical tools to predict what type of cafe (e.g., Coffee shop vs. Bakery vs. Restaurant) would be most suitable for a specific neighborhood based on market gaps and historical data.

d) High Failure Risks:

The inability to accurately assess the "co-location" patterns in which types of businesses thrive near each other increases the risk of business failure [2].

1.4 Project Objective

General Objective:

To develop a web-based cafe location recommendation system that utilizes GIS and Machine Learning to assist users in identifying the most suitable cafe type for a specific location in Kathmandu Metropolitan City.

Specific objectives:

- a) To develop an interactive web interface that allows users to pin specific locations in Kathmandu and visualize spatial data, including existing cafes and road networks, within a specified radius buffer.
- b) To analyze site suitability parameters by integrating diverse datasets (Google Maps, OpenStreetMap, and Population Census) to calculate competitor density, road accessibility, and business influence.
- c) To implement Machine Learning models that predicts the most suitable cafe type for a selected area based on identified market gaps and saturation levels.
- d) To provide actionable recommendations by displaying the top 5 existing cafes and suggesting one optimal new cafe type to open in that location, or the specified cafe type in another location.

1.5 Significance of the Study

The system reduces the uncertainty and financial risk associated with opening new businesses. By providing a data-backed recommendation (e.g., suggesting a "Bakery Cafe" where there is high footfall but low competition), it aids in strategic decision-making [1].

The project demonstrates the application of GIS in analyzing urban business distributions. It highlights how "co-location" patterns the tendency of certain businesses to cluster can be quantified and used for planning [2].

The project showcases the successful integration of web-based GIS visualization with backend Machine Learning logic. It contributes to the field of Location-Based Recommender Systems (LBRS) by moving beyond simple "nearest neighbor" searches to complex, multi-criteria suitability prediction [3].

CHAPTER 2

LITERATURE REVIEW

2.1 Related Works

The following table summarizes significant research studies related to location-based recommendation systems, spatial data mining, and multi-criteria decision-making (MCDM) for business suitability.

Table 1: Literature Review

Ref	Authors	Title	Methodology	Key Findings
[1]	V. Perez-Benitez, G. Gemar, and M. Hernández (2021)	Multi-Criteria Analysis for Business Location Decisions	Integrated PROMETHEE with AHP and GAIA analysis to rank 66 European cities.	Validated that security and technology are critical dimensions for business location decisions. Produced a preferential ranking of cities (e.g., London, Paris) for investment.
[2]	S. Han et al. (2024)	Restaurant location recommendation based on spatial data mining	Proposed a "LocationGCN" model combining spatial co-location pattern mining with GCN to capture spatial correlations of restaurants.	Demonstrated that different restaurant types have distinct co-location patterns. Achieved 74.88% accuracy in predicting suitable locations, outperforming traditional methods.
[3]	R. Sujithra and B. Surendiran (2020)	Location Based Recommendation Systems (LBRS) – A Review	A comprehensive review of existing LBRS services, categorizing them into geo-tagged media, point location, and Bayesian networks.	Identified that context-aware attributes (time, user preference) significantly improve recommendation quality. Bayesian networks are

			trajectory-based services.	noted for solving "cold start" problems.
[4]	M. Sarath, S. Saran, and K. V. Ramana (2018)	Site Suitability Analysis for Industries Using GIS MCDM	Used GIS and AHP to evaluate land suitability for industrial establishment based on six criteria (slope, and road proximity, land use, etc.) in Andhra Pradesh.	Proved that integrating GIS with MCDM provides a scientific basis for site selection, minimizing environmental damage, and optimizing resource use.
[5]	C. A. Onyekwe et al. (2024)	GIS-Based Site Suitability Study of Rice Farm Location	Employed GIS overlay analysis and reclassification of soil, slope, and accessibility data to map suitable farming sites in Nigeria.	Identified that 83.43% of the studied land was suitable for production. Demonstrated the efficacy of GIS in optimizing land resource utilization.

2.2 Summary of Literature Review

The review of existing literature highlights a significant evolution in location selection strategies, moving from manual, intuition-based decisions to sophisticated, data-driven approaches. Early methods and broad reviews, such as those by Sujithra & Surendiran [3], established the foundation of Location-Based Recommender Systems (LBRS), emphasizing the importance of geo-tagged data and user context. Building on this, studies like Perez-Benitez et al. [1] and Sarah et al. [4] demonstrated the power of integrating GIS with Multi-Criteria Decision Making (MCDM) techniques like AHP. These studies successfully validated that mathematical weighting of spatial factors (such as road accessibility and land use) yields highly accurate suitability maps, though they often focused on macro-level city selection or industrial zoning rather than small business needs.

More recent advancements, particularly by Han et al. [2], have introduced machine learning and "Prescriptive Analytics" into the domain. By utilizing advanced algorithms

like Graph Convolutional Networks, researchers have begun to model the complex "co-location" patterns of how specific businesses thrive when clustered near certain other types. However, a distinct gap remains for a localized, accessible tool for Kathmandu. Most existing systems are either too computationally complex for general users or lack the specific micro-level focus on cafe types (e.g., predicting "Bakery" vs. "Coffee Shop"). This project aims to bridge this gap by combining the proven GIS suitability analysis methods of [4] and [5] with a machine learning prediction model tailored to the specific urban dynamics of Kathmandu Metropolitan City.

2.3 Research Gap

Despite the extensive research in location suitability analysis, significant gaps remain when applying these methodologies to small-scale commercial enterprises like cafe in developing urban centers. Existing studies, such as the work by Perez-Benitez et al. (2021) and Sarath et al. (2018), predominantly focus on macro-level analysis ranking entire cities for investment or selecting large industrial zones which lacks the micro-level granularity required to evaluate specific street corners or neighborhoods for a coffee shop.

Furthermore, while recent advancements by Han et al. (2024) have introduced sophisticated deep learning models for location prediction, these systems are often computationally intensive and focus on binary suitability (whether a location is "good" or "bad") rather than recommending the optimal business type (e.g., Bakery vs. Cafe) from multiple options.

Additionally, traditional GIS approaches, as seen in Onyekwelu et al. (2024), are frequently tailored to agriculture, relying on physical factors like soil and slope while neglecting dynamic socio-economic indicators such as competitor density and market saturation that are critical for the hospitality industry.

Finally, comprehensive reviews like those by Sujithra & Surendiran (2020) provide theoretical frameworks but highlight a lack of practical, hybrid systems that effectively solve the "cold start" problem for new businesses without historical data. This project aims to address these limitations by developing a specialized, web-based tool that combines micro-level spatial analysis with machine learning to provide accessible, type-specific recommendations for Kathmandu's cafe entrepreneurs.

CHAPTER 3

REQUIREMENT ANALYSIS

3.1 Functional Requirements

Functional requirements define the specific behaviors or functions the system must perform. Based on the project roadmap, the key functional requirements include:

- a) User Authentication: The system shall allow users to sign in securely using existing Google accounts to maintain a personalized profile of saved locations.
- b) Interactive Map Visualization: The system shall provide a high-resolution map of Kathmandu Metropolitan City with zoom-in/out capabilities using OpenStreetMap and Google Maps API.
- c) Location Pinning: Users shall be able to pin a specific geographic location on the map to define the center of their business interest.
- d) Buffer Analysis (Specified Radius): The system shall automatically calculate and visualize a specified radius buffer around the pinned location to analyze the "business influence area".
- e) Spatial Data Retrieval: The system shall fetch and display existing cafes, road networks, and population density data within the selected buffer.
- f) Recommendation Engine:
 - i. Existing Success: The system shall list the Top 5 existing cafes in the area based on popularity and ratings.
 - ii. Predictive Suggestion: The system shall utilize a Random Forest algorithm (with others if it fails) to predict and recommend one specific cafe type (e.g., Bakery, Coffee Shop) that is most suitable for the location based on market gaps.
- g) Suitability Parameter Display: The system shall display a dashboard showing the quantitative values for parameters like competitor density, accessibility, and demographic suitability.

3.2 Non-Functional Requirements

Non-functional requirements specify the criteria that can be used to judge the operation of a system rather than specific behaviors.

- a) Performance: The spatial analysis and ML prediction should be processed within 3 to 5 seconds to ensure a smooth user experience.
- b) Usability: The web interface must be intuitive, requiring minimal technical knowledge for an entrepreneur to navigate the map and interpret results.
- c) Reliability: The system should provide consistent recommendations based on the underlying datasets from Google Maps and OSM.
- d) Scalability: The architecture should be designed to allow for the inclusion of other Metropolitan Cities beyond Kathmandu or additional business categories in the future.
- e) Accuracy: The ML model should aim for high precision in its predictions, validated against historical business success patterns in the urban context [2].

3.3 System Requirements

To ensure the effective development and deployment of the cafe recommendation system, the following hardware and software specifications are required:

3.3.1 Hardware Requirements

- a) Processor: Intel i5 or higher (for handling local ML model training and spatial data processing).
- b) RAM: Minimum 8GB (to handle GIS rendering and data manipulation).
- c) Storage: 20GB of free space (for local datasets and environment setup).

3.3.2 Software Requirements

- a) Operating System: Windows 10/11, macOS, or Linux.
- b) Frontend: HTML/CSS/JavaScript with Leaflet.js for map rendering.
- c) Backend: Python (Django) to handle GIS logic and ML integration.
- d) Database: PostgreSQL with PostGIS extension (for spatial queries).
- e) Machine Learning: Scikit-learn library for the Random Forest algorithm.
- f) Data Sources: Google Places API, OpenStreetMap data, and National Population Census datasets.

3.4 Feasibility Analysis

3.4.1 Technical Feasibility

The project is technically feasible as the tools required Python for ML, PostGIS for spatial data, and HTML, CSS, JS for the frontend are well-documented and widely used. The availability of APIs like Google Maps and open data from

OSM ensures that the necessary spatial datasets can be retrieved and integrated effectively.

3.4.2 Economic Feasibility

The project is economically feasible because it primarily utilizes open-source technologies (Python, Leaflet, PostgreSQL). The primary costs involved are related to API usage (Google Cloud) and hosting, which can be managed within a student project budget or via free-tier credits. For the end-user, this system provides high value by reducing the financial risk of business failure [1].

3.4.3 Operational Feasibility

The system is highly operational as it solves a real-world problem for entrepreneurs in Kathmandu. By automating the complex process of spatial co-location mining and market gap analysis tasks that currently require expensive consultants the platform offers a practical solution that users can easily adopt into their business planning workflow [2].

CHAPTER 4

SYSTEM DESIGN AND ARCHITECTURE

4.1 System Architecture Design

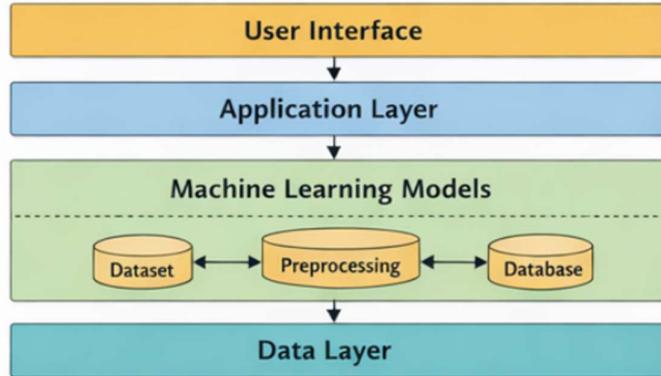


Figure 1: Design of System Architecture

This diagram illustrates the structural layers of the map-based cafe location recommendation system for Kathmandu Metropolitan City. The User Interface and Application Layer represent the web-based platform where users sign in via Google, interact with the Kathmandu Metropolitan City map, select cafe types, and pin locations. The Machine Learning Models layer utilizes algorithms like Random Forest to process spatial and historical data including market gaps and competitor density to predict ideal business types. Supporting this is the Data Layer, which integrates the project's multi-source datasets from Google Maps, OpenStreetMap, and Population Census records to determine site suitability within a 500m radius.

4.1.1 User Interface Layer

The User Browser is the platform through which the user accesses the system using a web browser such as Chrome or Edge. It allows the user to open the website, sign in with a Google account, and interact with the map-based features of the application. The Web Interface provides the visual and interactive elements of the system. It includes the login page, cafe selection options, buttons, forms, and result panels. This interface captures user inputs such as cafe type selection, map pin location, and recommendation requests, and sends them to the application layer for processing. The Interactive Kathmandu Map displays a zoomable and clickable map of Kathmandu Metropolitan

City. It allows users to pin a location, view a 500-meter radius area (default), see existing cafes, and visualize suitability using heatmaps. This component enables intuitive location-based interaction and presents analysis results in a user-friendly manner.



Figure 2: User Interface Layer

4.1.2 Application Layer

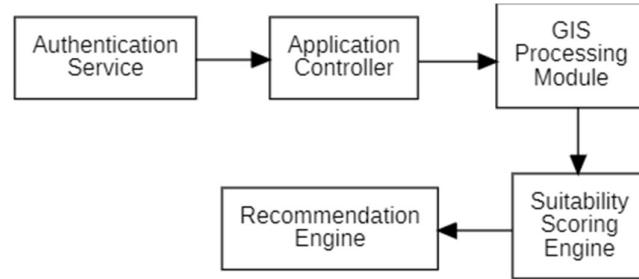


Figure 3: Application Layer

Google Authentication Service allows users to securely sign in using their existing Google account. It ensures authorized access to the system before users can interact with the Kathmandu Metropolitan City map and cafe analysis features. Application Controller manages all user requests such as cafe selection, map pinning, and recommendation requests. It controls the workflow and coordinates communication between different system modules. GIS Processing Module processes the pinned map location and creates a 500-meter analysis radius. It extracts spatial data such as existing cafes, roads, and nearby facilities using Google Maps, OpenStreetMap, and census data. Suitability Scoring engine evaluates how suitable the selected location is for the chosen cafe type. It analyzes factors like competition, population density, accessibility, and demand to calculate a suitability score. The Recommendation Engine provides the final results by suggesting the top five existing cafes in the area and predicting one new cafe type that can be successfully opened within the selected radius.

4.1.3 Machine Learning Layer

The Machine Learning layer serves as the core analytical engine of the system, responsible for processing complex spatial datasets and generating predictive insights for cafe site selection. This layer begins with a robust feature engineering process where

raw data including competitor density within a 500-meter buffer, accessibility indices from OpenStreetMap, and demographic statistics from the 2021 Census are converted into numerical vectors. At the heart of this layer lies the Random Forest Classifier, an ensemble learning model that utilizes a collection of decision trees to evaluate these spatial features. During the training phase, the model identifies non-linear patterns from historical cafe performance data in Kathmandu, learning to differentiate between optimal and sub-optimal business locations. When a user interacts with the web-based GIS interface, the ML layer performs real-time inference by passing the selected coordinate spatial attributes through its trained forest of trees. Each tree contributes a vote toward a final suitability classification, ensuring that the recommendation is backed by a statistical consensus rather than a single variable. Finally, the layer outputs a suitability score which is passed back to the frontend for visualization, effectively bridging the gap between raw geographical data and prescriptive business decision-making.

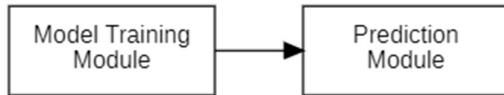


Figure 4: Machine Learning Layer

4.1.4 Data Layer

The Data Layer forms the backbone of the proposed location-based cafe recommendation system by providing reliable and structured data for analysis and prediction. This layer integrates multiple real-world data sources into a centralized database, ensuring consistency and efficient data access across the system. All collected datasets are preprocessed to remove duplicates, handle missing values, and standardize geographic coordinates before storage.

The Google Maps dataset supplies detailed information about existing cafe in Kathmandu Metropolitan City, including their geographic locations, cafe categories, user ratings, and review counts. This data is essential for identifying competitors, analyzing cafe popularity, and ranking existing cafe within the selected 500-meter radius.

The OpenStreetMap dataset provides geospatial and infrastructure-related information such as road networks, nearby facilities, and land-use patterns. These features help

evaluate accessibility, connectivity, and environmental suitability of a location, which directly influence cafe success.

The population census dataset contributes to demographic information such as ward-wise population density and distribution. This data is used to estimate potential customer demand and footfall in the selected area. By combining these datasets within a central database, the Data Layer ensures accurate spatial analysis, effective suitability scoring, and reliable machine learning-based prediction of new cafe opportunities.

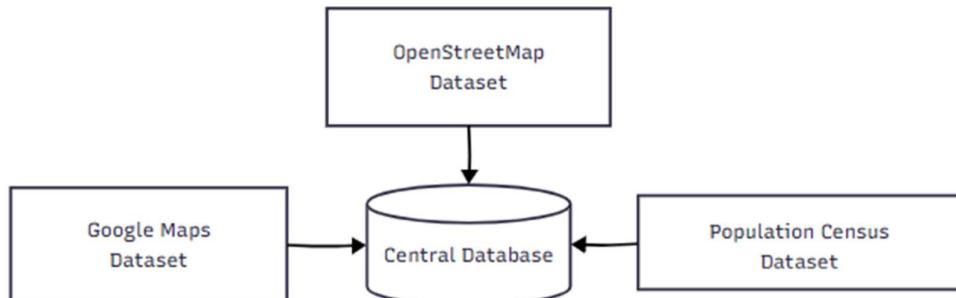


Figure 5: Data Layer

4.2 Use Case Diagram

The system involves three main actors: the User, who wishes to open a cafe and interacts with the application to view maps, analyze locations, and receive recommendations; the Admin, who manages datasets such as census and map data and retrains the machine learning model to maintain accuracy; and the Map Service (OpenStreetMap/Google Maps), which provides the base map of Kathmandu Metropolitan City with zoom and pan functionalities. The user logs in using a Google account, views the interactive Kathmandu map, selects a cafe type, and pins a location to define the analysis area. Based on the pinned point, the system displays nearby cafes and suitability factors such as population, road access, and cafe ratings within the specified radius. Using this data, the system recommends the top existing cafes and predicts the most suitable type of new cafe to open in that location, while the admin continuously updates datasets and the ML model to ensure reliable and accurate recommendations.

4.3 Data Flow Diagram (DFD)

It provides a way to visualize how data enters and leaves the system, what changes the data, and where data is stored.

DFD Level 0

It defines the scope of the system by identifying the boundaries between the system and external entities, such as the User and external Data Sources (Google Maps, OSM, and Census Data). In this system, the user provides inputs like cafe type, and the system provides recommendations and suitability analysis as output.

DFD Level 1

It illustrates how user inputs are processed through various stages: authentication, data retrieval, spatial analysis, and machine learning prediction. This level highlights the interaction between functional components and internal data stores like the spatial database.

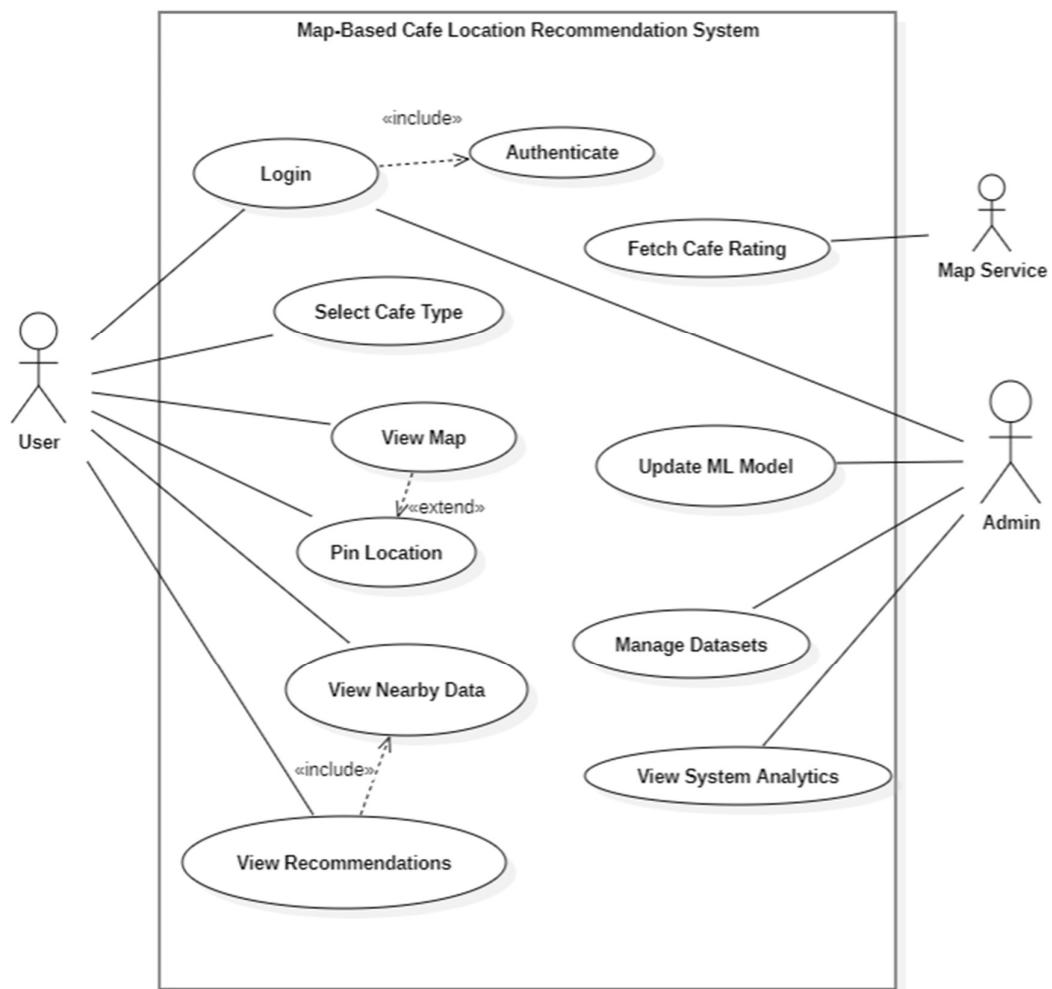


Figure 6: Use Case Diagram for Overall Operation

4.4 Flowchart of the System

The process begins with user login, followed by loading a map of Kathmandu Metropolitan City where users select their preferred cafe type and pin their desired location on the map. The system then analyzes a 500-meter radius (specified) around the pinned location and calculates the suitability of that area for opening a cafe based on various factors. If the location is deemed suitable, the system proceeds to recommend the top 5 cafes in that area and uses predictive analytics to determine which cafe type would be best to open there, finally displaying these results to the user. If the location is not suitable, system will automatically suggest the best location, and also users can return to pin a different location and repeat the analysis process until they find an appropriate spot for their cafe business.

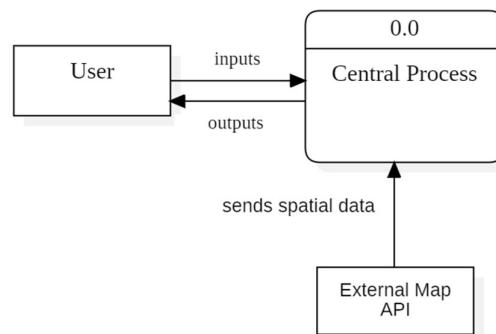


Figure 7: DFD Level 0

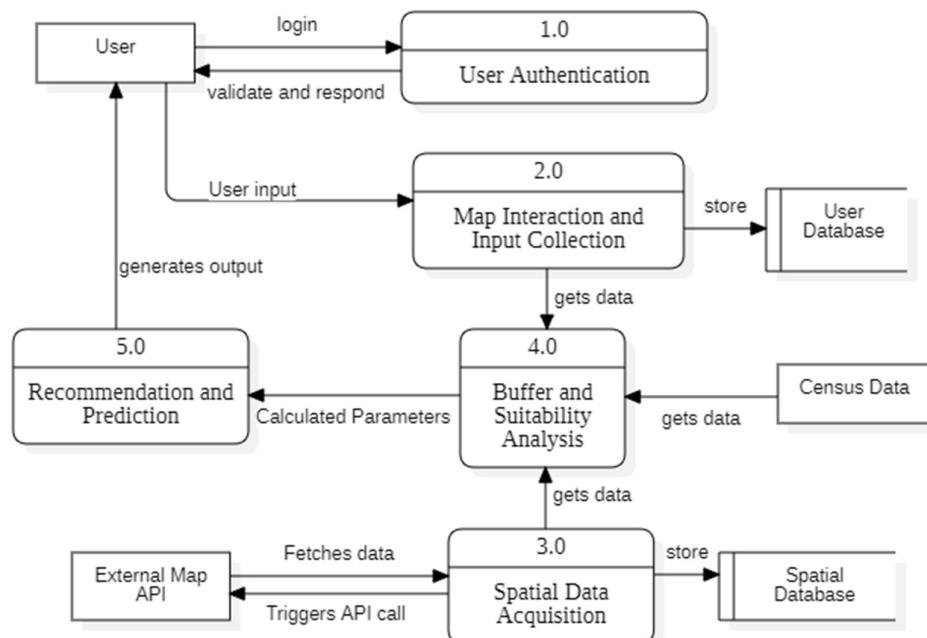


Figure 8: DFD Level 1

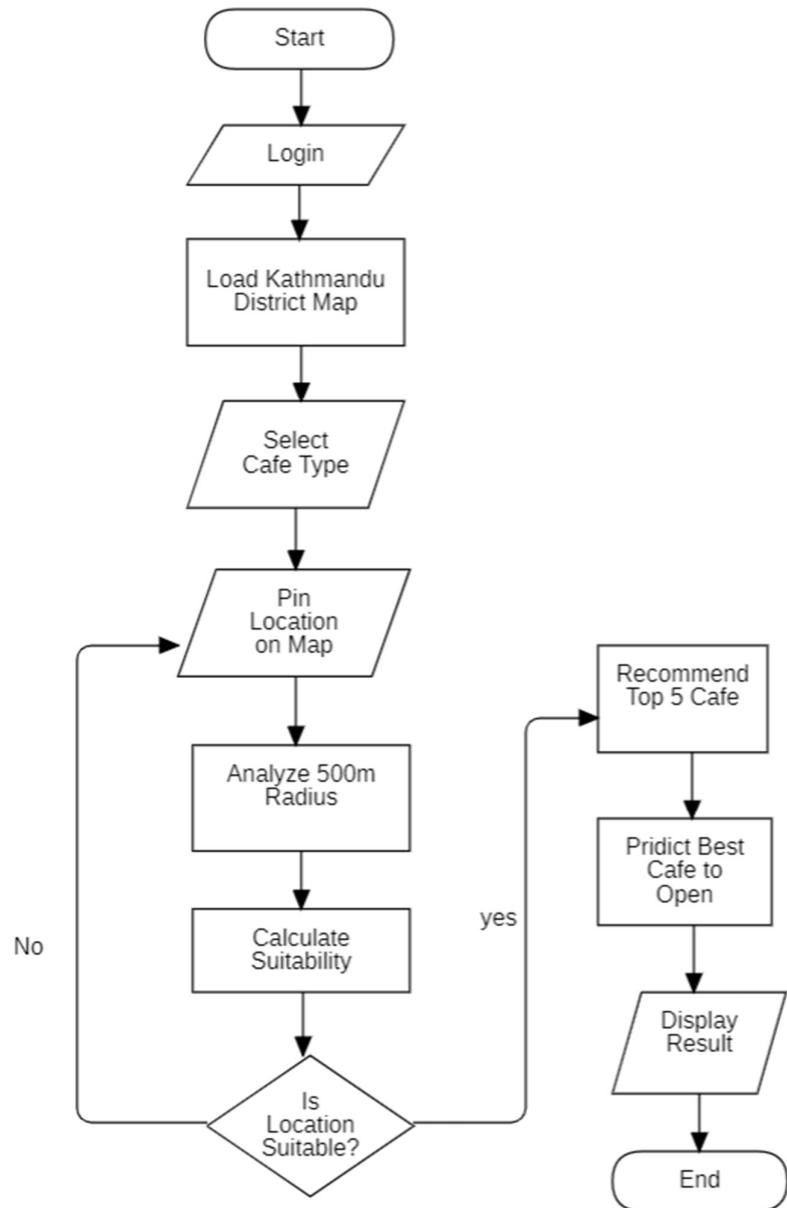


Figure 9: Flowchart of the System

CHAPTER 5

METHODOLOGY

This chapter details the technical steps, the mathematical foundations of the predictive model, and the software development lifecycle adopted for this project.

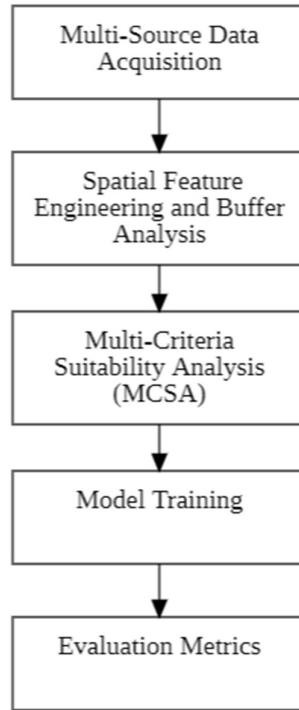


Figure 10: Methodology Overview

5.1 Multi-Source Data Acquisition

To build a comprehensive recommendation engine, data is harvested from three heterogeneous sources:

- a) Google Places API: This is the primary source for identifying current market players. It provides the most "real-world" and updated dataset for existing businesses.
 - a. Collection Method: We utilize the Places API via HTTP GET requests. For a specific pinned location in Kathmandu, the system triggers a nearby search or search text query.
 - b. Parameters used:
 - i. location: Latitude and longitude of the pinned point.
 - ii. radius: 500 meters (adjustable).

- iii. type: Filtered for cafe, bakery, and restaurant.
- c. Extracted Attributes: Business Name, Place ID, Global Rating, Total User Ratings, and current operational status.
- d. Validity Check: Google data is validated by millions of users; however, to ensure "freshness," we only extract businesses with a business status labeled as OPERATIONAL.
- b) OpenStreetMap (OSM): Provides the structural framework for the analysis. We extract the "Highway" and "Amenity" layers to analyze road connectivity and existing urban infrastructure. Data is mined using the Overpass Turbo API. We execute a specialized query to extract all "Highways" (roads) and "Amenity" points within the Kathmandu Metropolitan City boundary.
- c) National Population Census (2021): Provides the demographic weight. Population density data is linked to ward-level shapefiles to estimate the potential customer volume in specific neighborhoods.

To understand the "potential customer base," we integrate the latest official statistics from the National Statistics Office (NSO) of Nepal.

- a) Collection Method: Data is sourced from the Census Nepal 2021 Data Portal. We extract ward-level CSV files for Kathmandu Metropolitan City.
- b) Key Metrics: Population density (people per sq. km) and household counts at the ward level.
- c) Data Linking: Since census data is tabular, we perform a Spatial Join in the backend, linking the tabular population counts to a GIS Shapefile of Kathmandu's administrative wards.

5.2 Spatial Feature Engineering and Buffer Analysis

Once the data is cleaned and standardized to the WGS84 coordinate system, the system performs "Buffer Analysis."

- a) Centroid Selection: When a user clicks on the map, the system captures the latitude and longitude (x, y).
- b) 500m Influence Zone: A circular buffer with a radius of 500 meters is generated around the pin. This distance is chosen based on urban "walkability" standards for food services.

- c) Feature Extraction: The system counts all existing cafes (Competitor Density), measures total road length (Accessibility Index), and extracts ward-level density (Demographic Demand) within this buffer.

5.3 Multi-Criteria Suitability Analysis (MCSA)

Following the methodology of Sarath et al. (2018), the system applies weighted scoring to these spatial features. This transforms raw counts into a suitability score. If a location has low competitor density but high population density and road accessibility, it receives a high "Market Opportunity" rank.

5.4 Model Training

The core of the system's "Prescriptive Analytics" is the Random Forest Classifier. Unlike simple search tools, it predicts the optimal type of cafe for a given location based on patterns learned from successful businesses.

5.4.1 Random Forest:

Random Forest is an ensemble learning method that constructs a multitude of Decision Trees during training and outputs the class that is the mode of the classes of the individual trees.

The model chooses the best "split" at each node of a decision tree using Gini Impurity or Entropy.

- a) Gini Impurity (G): Measures the frequency at which a randomly chosen element from the set would be incorrectly labeled.

(Where P_i is the probability of an item belonging to class i).

- b) Entropy (H): Measures the level of impurity or "randomness" in the data.

- c) Information Gain (IG): The model calculates which feature (e.g., population vs. road density) provides the most information to reduce entropy.

$$IG = H(Parent) - \sum \left(\frac{N_{child}}{N_{parent}} \times H(child) \right) \dots \dots \dots (3)$$

5.4.2 Training Process

- a) Bootstrap Aggregating (Bagging): The model creates multiple subsets of the training data with replacement.

- b) Feature Randomness: In each tree, only a random subset of spatial features (e.g., just road density and competitors) is considered at each split to ensure diversity and prevent overfitting.
- c) Voting: The final recommendation (e.g., "Bakery") is the one predicted by the majority of the individual trees.

5.5 Evaluation Metrics

To ensure the recommendation system is reliable, the Random Forest model is evaluated using a Confusion Matrix to compare predicted vs. actual successful business types in Kathmandu.

Table 2: Evaluation Metrics

Metric	Formula	Description
Accuracy	$\frac{TP + TN}{TP + TN + FP + FN}$	Overall percentage of correct predictions.
Precision	$\frac{TP}{TP + FP}$	Measures the quality of the positive predictions (how many predicted "Bakeries" were suitable).
Recall	$\frac{TP}{TP + FN}$	Measures the ability to find all suitable locations in the dataset.
F1-Score	$2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$	The harmonic mean of precision and recall, providing a single score for model performance.

5.6 Software Development Model: Agile Model

For this project, the Agile Development Model is adopted due to its iterative nature, which is ideal for refining complex ML models and GIS interfaces.

Phase 1: Planning and Backlog

In this phase, we define the "Product Backlog," which includes setting up the PostgreSQL/PostGIS database and integrating the Google Maps API.

Phase 2: Design and Prototyping

The UI/UX is designed using React.js. An initial prototype of the interactive Kathmandu map is built, allowing for simple coordinate pinning.

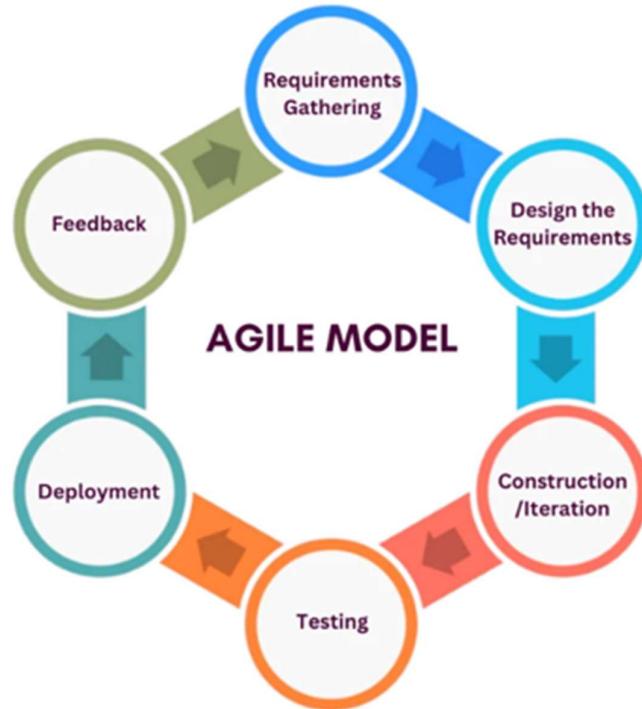


Figure 11: Agile Software Development Model

Source: <https://medium.com/tech-stack-insights/the-software-development-lifecycle-sdlc-a-deep-dive-with-agile-methodology-f7846853b7a3>

Phase 3: Development and Iteration

This is the core "Build" phase. The Random Forest model is trained on Kathmandu business data. The backend (Django/Flask) is developed to handle the spatial queries between the frontend map and the ML engine.

Phase 4: Testing & Review

The system undergoes unit testing for the spatial buffer logic and accuracy testing for the recommendation engine. Feedback is used to "tune" the hyperparameters of the Random Forest model to improve the F1-score.

CHAPTER 6

EXPECTED OUTPUT

The proposed system aims to deliver a functional web-based Geographic Information System (GIS) integrated with Machine Learning to assist entrepreneurs in Kathmandu. The expected outcomes of this project are divided into visual, analytical, and predictive outputs.

6.1 Interactive Web Interface

The primary output will be a platform-independent web application featuring:

- a) Authentication Module: A secure login system utilizing Google OAuth 2.0.
- b) Geospatial Dashboard: An interactive map of Kathmandu Metropolitan City (integrated via OpenStreetMap/Leaflet) allowing users to pan, zoom, and drop pins on specific coordinates.
- c) Parameter Visualization: A dashboard overlay showing real-time data for a 500m radius, including population density, road accessibility, and existing competitor locations.

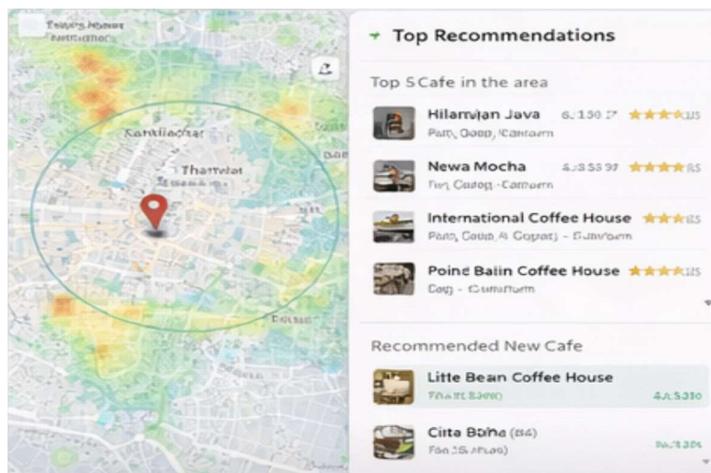


Figure 12: Expected Output Web Interface

6.2 Spatial Analysis Results

Upon pinning a location, the system will output a detailed spatial report for the 500m buffer zone:

- a) Competitor Mapping: Visualization of all existing cafes within the radius.

- b) Suitability Metrics: A summary of collected datasets (Census, OSM, Google Maps) presented in charts or heatmaps to show why a location is or isn't suitable for the selected business type.

6.3 Machine Learning Based Recommendations

The core innovation of the project is the dual-layer recommendation output generated by the Random Forest model:

Top 5 Existing Cafes

The system will identify and list the five most successful existing cafes within the 500m radius. These are ranked based on:

- a) Average user ratings and review counts.
- b) Popularity and estimated footfall indicators.
- c) Location-specific advantages.

New Business Prediction (The Market Gap)

The system will provide a specific prediction for a new cafe type that has the highest potential for success at the pinned location.

- a) Market Gap Analysis: Identification of cafe types that are missing in the area despite high population demand.
- b) Saturation Logic: If the user's selected cafe type is "over-saturated," the system will suggest an alternative type (e.g., suggesting a "Bakery" if "Coffee Shops" are too numerous).

CHAPTER 7

TIME SCHEDULE

7.1 Project Timeline Overview

The "Map-Based Cafe Location Recommendation System" will be developed over a period of 8 weeks (2 months) following the Agile Software Development methodology. The timeline is designed to be intensive and iterative, allowing for the parallel development of the map interface and the machine learning model to ensure timely completion.

7.2 Gantt Chart

The following Gantt chart illustrates the weekly distribution of tasks over the 8-week duration.

Table 3: Gantt Chart

Major Steps	W1	W2	W3	W4	W5	W6	W7	W8
Planning and Requirements								
Data Collection and Design								
Web Interface Dev								
ML Model Training								
Integration and Testing								
Deployment								
Documentation								

7.3 Detailed Task Breakdown

The following Gantt chart illustrates the weekly distribution of tasks throughout the project lifecycle.

The project follows agile model development approach. It begins with Planning and Requirements analysis in the first two weeks to define the system's scope and finalize the technical stack for GIS and Machine Learning. Overlapping with the second week, Data Collection and Design involves gathering spatial datasets from OpenStreetMap and census data, which provides the "raw material" needed for the subsequent phases

As the data becomes available, Web Interface Development kicks off in week three to build the React map interface. To ensure efficiency, ML Model Training begins in week four, allowing to train the Random Forest algorithm on the data collected earlier. Once the interface and the "brain" (the model) are ready, Integration and Testing take place in weeks six and seven to ensure the recommendation logic displays correctly on the map. The final two weeks are dedicated to Deployment (setting up the live environment) and Documentation, ensuring all technical findings and user manuals are compiled into the final report for the Week 8 deadline.

REFERENCES

- [1] V. Perez-Benitez, G. Gemar, and M. Hernández, "Multi-Criteria Analysis for Business Location Decisions," *Mathematics*, vol. 9, no. 2615, 2021.
- [2] S. Han, L. Chen, Z. Su, S. Gupta, and U. Sivarajah, "Identifying a good business location using prescriptive analytics: Restaurant location recommendation based on spatial data mining," *Journal of Business Research*, vol. 179, p. 114691, 2024.
- [3] R. Sujithra @ Kanmani and B. Surendiran, "Location Based Recommender Systems (LBRS) – A Review," in *International Conference on Computational Intelligence and Data Science (ICCID 2020)*, 2020.
- [4] M. Sarath, S. Saran, and K. V. Ramana, "Site Suitability Analysis for Industries Using GIS and Multi Criteria Decision Making," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. IV-5, pp. 447–454, 2018.
- [5] C. A. Onyekwelu et al., "GIS-Based Site Suitability Study of Rice Farm Location in Bende Local Government Area, Abia State, Nigeria," *African Journal of Food, Agriculture, Nutrition and Development*, vol. 24, no. 5, pp. 26312–26332, May 2024.