IMPLEMENTATION OF GEOLOCATION EXTRACTION DATA ANALYTICS USING ML

A MINOR PROJECT - III REPORT

Submitted by

KESAVI R	**927621BEC087**
KIRUTHIKA V R	**927621BEC091**
MONIKA G	**927621BEC127**

BACHELOR OF ENGINEERING

in

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

M.KUMARASAMY COLLEGE OF ENGINEERING

(Autonomous)

KARUR – 639 113

OCTOBER 2023

M.KUMARASAMY COLLEGE OF ENGINEERING, KARUR

BONAFIDE CERTIFICATE

Certified that this **18ECP105L - Minor Project - III** report **“IMPLEMENTATION OF GEOLOCATION EXTRACTION DATA ANALYTICS USING ML”** is the bonafide work of “KESAVI.R (927621BEC087), KIRUTHIKA.V.R (927621BEC091), MONIKA.G (927621BEC127) “ who carried out the project work under my supervision in the academic year **2023 -2024 - ODD SEMESTER**.

SIGNATURE	**SIGNATURE**
Dr.A.KAVITHA B.E., M.E., Ph.D.,	Mrs.S.SIVARANJANI B.E.,M.E.,
HEAD OF THE DEPARTMENT,	**SUPERVISOR,**
Professor,	Assistant Professor,
Department of Electronics and	Department of Electronics and
Communication Engineering,	Communication Engineering,
M.Kumarasamy College of Engineering,	M.Kumarasamy College of Engineering,
Thalavapalayam,	Thalavapalayam,
Karur-639113.	Karur-639113.

This report has been submitted for the **18ECP105L – Minor Project - III** final review held at M.Kumarasamy College of Engineering, Karur on _____________________

PROJECT COORDINATOR

INSTITUTION VISION AND MISSION

Vision

To emerge as a leader among the top institutions in the field of technical education.

Mission

M1: Produce smart technocrats with empirical knowledge who can surmount the global challenges.

M2: Create a diverse, fully -engaged, learner -centric campus environment to provide quality education to the students.

M3: Maintain mutually beneficial partnerships with our alumni, industry and professional associations

DEPARTMENT VISION, MISSION, PEO, PO AND PSO

Vision

To empower the Electronics and Communication Engineering students with emerging technologies, professionalism, innovative research and social responsibility.

Mission

M1: Attain the academic excellence through innovative teaching learning process, research areas & laboratories and Consultancy projects.

M2: Inculcate the students in problem solving and lifelong learning ability.

M3: Provide entrepreneurial skills and leadership qualities.

M4: Render the technical knowledge and skills of faculty members.

Program Educational Objectives

PEO1: **Core Competence:** Graduates will have a successful career in academia or industry associated with Electronics and Communication Engineering

PEO2: **Professionalism:** Graduates will provide feasible solutions for the challenging problems through comprehensive research and innovation in the allied areas of Electronics and Communication Engineering.

PEO3: **Lifelong Learning:** Graduates will contribute to the social needs through lifelong learning, practicing professional ethics and leadership quality

Program Outcomes

PO 1: Engineering knowledge: Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

PO 2: Problem analysis: Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences.

PO 3: Design/development of solutions: Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations.

PO 4: Conduct investigations of complex problems: Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions.

PO 5: Modern tool usage: Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations.

PO 6: The engineer and society: Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice.

PO 7: Environment and sustainability: Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development.

PO 8: Ethics: Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.

PO 9: Individual and team work: Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.

PO 10: Communication: Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.

PO 11: Project management and finance: Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.

PO 12: Life-long learning: Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

Program Specific Outcomes

PSO1: Applying knowledge in various areas, like Electronics, Communications, Signal processing, VLSI, Embedded systems etc., in the design and implementation of Engineering application.

PSO2: Able to solve complex problems in Electronics and Communication Engineering with analytical and managerial skills either independently or in team using latest hardware and software tools to fulfil the industrial expectations.

Abstract	**Matching with POs, PSOs**
Suitable accommodation, Weather condition, Proximity to essential locations, Daily requirements	PO1, PO2, PO4, PO6, PO7, PO8, PO9, PO10, PO11, PO12, PSO1, PSO2

ACKNOWLEDGEMENT

Our sincere thanks to **Thiru.M.Kumarasamy, Chairman** and **Dr.K.Ramakrishnan, Secretary** of **M.Kumarasamy College of Engineering** for providing extraordinary infrastructure, which helped us to complete this project in time.

It is a great privilege for us to express our gratitude to **Dr.B.S.Murugan., B.Tech., M.Tech., Ph.D., Principal** for providing us right ambiance to carry out this project work.

We would like to thank **Dr.A.Kavitha, B.E., M.E., Ph.D., Professor and Head**, **Department of Electronics and Communication Engineering** for her unwavering moral support and constant encouragement towards the completion of this project work.

We offer our wholehearted thanks to our **Project Supervisor, Mrs.S.SIVARANJANI B.E.,M.E., Assistant Professor**, Department of Electronics and Communication Engineering for her precious guidance, tremendous supervision, kind cooperation, valuable suggestions, and support rendered in making our project successful.

We would like to thank our **Minor Project Co-ordinator**, **Dr.K.Karthikeyan, B.E., M.Tech., Ph.D., Associate Professor**, Department of Electronics and Communication Engineering for his kind cooperation and culminating in the successful completion of this project work. We are glad to thank all the Faculty Members of the Department of Electronics and Communication Engineering for extending a warm helping hand and valuable suggestions throughout the project. Words are boundless to thank our Parents and Friends for their motivation to complete this project successfully.

ABSTRACT

Embarking on a new journey, whether for education or work, marks the path to self-discovery and self-reliance. This venture is not without its challenges, including navigating unfamiliar healthcare systems, ensuring personal safety, and managing finances. Yet, the most prominent challenge often lies in securing suitable accommodation. Students and young professionals frequently encounter hurdles when relocating to new cities or states for higher education or employment. Each individual possesses unique priorities and preferences, making it a daunting task to find affordable housing with easy access to daily necessities. This project's core objective is to develop a user-centric system that identifies the ideal accommodation in a given city. Users can specify their criteria, such as weather condition, proximity to essential locations, and daily requirements, enabling the system to recommend the most fitting options. Furthermore, this versatile system can extend its utility to other domains, such as assisting businesses in locating optimal sites (e.g., positioning restaurants or cafes near educational institutions) or identifying prime agricultural land for maximum crop yield.

Keywords: Machine learning,Weather condition, Proximity to essential locations, Daily requirements, Python,

TABLE OF CONTENTS

CHAPTER No.	CONTENTS		PAGE No.
	Institution Vision and Mission		iii
	Department Vision and Mission		iii
	Department PEOs, POs and PSOs		Iv
	Abstract		viii
	List of Tables		xi
	List of Figures		xii
	List of Abbreviations		xiii
1	**INTRODUCTION**		**1**
	1.1	Problem statement	2
	1.2	Objectives	2
2	**LITERATURE SURVEY**		**3**
	2.1	Implementation of GPS for location tracking	3
	2.2	Geolocation analysis using machine learning	3
3	**EXISTING SYSTEM`**		**5**
4	**PROPOSED SYSTEM**		7
5	**SOFTWARE USED**		**9**
6	**SOURCE CODE**		**15**
7	**RESULT AND DISCUSSION**		21
8	**CONCLUSION AND FUTUTE WORK**		**23**
	REFERENCES		**24**

LIST OF FIGURES

FIGURE No.	TITLE	PAGE No.
7.1	**Details of pincode**	**21**
7.2	**Details of latitude and longitude**	**21**
7.3	**Details of weather conditions**	**22**

LIST OF ABBREVIATIONS

ACRONYM		ABBREVIATION
API	-	Application programming interface
HTTP	-	Hyper Text Transfer Protocol
XML	-	Extensible Markup Language

CHAPTER 1

INTRODUCTION

Geolocation extraction using machine learning is a powerful and increasingly essential technology for a wide range of applications. This innovative approach combines the capabilities of machine learning algorithms with geospatial data to accurately determine the geographic coordinates of a given location, such as an address or point of interest. By harnessing the power of ML, we can automate the process of geolocation extraction, making it faster, more accurate, and scalable. This technology has far-reaching implications, from enhancing location-based services in mobile apps to improving logistics and supply chain management. In this discussion, we'll explore the key concepts, methods, and challenges involved in implementing geolocation extraction using machine learning, and delve into the exciting possibilities it presents for various industries. Geolocation technology is increasingly used in our daily lives. From navigation apps to social media platforms, geolocation technology is used to pinpoint our location and deliver personalized services. It has numerous applications across various industries, and it's becoming increasingly important for businesses to leverage its capabilities. Geolocation technology is used for navigation and location-based services such as finding directions, discovering nearby restaurants, and finding a service. Geolocation data is becoming increasingly important for e-commerce companies, as it allows them to provide more personalized and targeted experiences for their customers. For example, an online retailer can use geolocation data to offer promotions or discounts to customers who are located near a particular store. IP Geolocation is a web-based service that allows you to determine an IP addresses proximate geographical location (usually with city-level accuracy). Generally, it's a simple HTTP call that can be used across different platforms and programming languages.

1.1. PROBLEM STATEMENT

Relying solely on a pincode is insufficient for determining the exact geographic location and its corresponding weather conditions. While pincodes provide a basic level of geographic information, they typically encompass a broader area, such as a neighborhood or town, and do not account for variations in weather that can occur within that larger region. To obtain accurate weather data, it is essential to use more specific geographical coordinates, such as latitude and longitude, which can pinpoint an exact location and provide precise weather information.

1.2. OBJECTIVES

Understanding that our system will be used in critical applications where accuracy matter a lot, we have objectives as to analyse and verify the exact location and its weather condition using zip code.

CHAPTER 2

LITERATURE SURVEY

Recommendation systems have come into play to present the user with exactly what he had in mind while searching for an item or service. The system takes into consideration the parameters input by the user so as to determine what he actually needs. Today we are surrounded by different types of applications and websites for fulfilling our daily needs and all these technologies are using the recommendation system in some way or the other, for example recommending similar types of music based on your previously created playlist or songs that you frequently listen, similar genres of movies that have been previously watched or searched for, and same goes for the books. The amount of information available on the internet is increasing with every tick of the clock hence providing the user a plethora of options to choose from, that's where the recommendation system comes in to help decide what to choose by taking account of the user's history and relevant suggestions, a bulk of information gets eliminated to present a few similar options. Various studies and researches on decision-making strategies were done to analyze the methods that can be relied upon for efficient decision-making that helped many planners, students, and practitioners interested in the field of spatial location decision-making. Huff model is a type of spatial interaction model that measures the probabilities of customers based on gravity at every location. Multiple regression analysis is a technique that implies factors affecting the sales of current stores would have a similar impact on stores located at another location, it is used mostly by restaurant chains, books, music, and home furnishing stores. The analog approach takes into consideration the aspects of trade and site area features to recommend a similar site. The gravity model is used at 2 ends of the spectrum, where the first spectrum considers the number of sales and probability of shoppers and the second spectrum focuses on distance as a major factor for site location. Neural network is a computational model that simulates functional aspects and processes information with help of a connectionist approach and is considered as a robust classifier. Machine learning is mainly used for traditional techniques like geostatistics and is widely used for spatial application and GIS. Multi-criteria decision-making model is considered as an important tool to solve complex business solutions as they tend to deal with uncertainty, complexity, and conflicting objectives.

TITLE: Implementation of GPS for location tracking

AUTHOR: Ahmad Ashraff Bin Ariffin

PUBLICATION: Control and System Graduate Research Colloquium (ICSGRC), 2011 IEEE

TITLE: Geolocation Analysis Using Machine Learning

AUTHOR: Sakshi Rajesh Sinha, Prof. Sumedh Pundkar

PUBLICATION: International Journal of Engineering Research in Computer Science and Engineering (IJERCSE), June 2022

CHAPTER 3
EXISTING SYSTEM

Using a pin code allows us to determine the general geographic location of a specific area. Pin codes are postal codes or ZIP codes that help organize and route mail and parcels efficiently. Each pin code typically represents a particular region, locality, or district. However, it's important to note that pin codes alone do not provide finer details about a location, such as its exact coordinates or more specific geographical features. To obtain precise location information, additional methods, such as GPS coordinates or detailed address information, are necessary.

The first digit of a PIN code designates the zone, the second digit is a subzone, the third digit implies the sorting district within that region, and the last three digits indicate the specific post office within that district.

The PIN code system was introduced in the country to assist the postal department in manually sorting and delivering mail to the correct location. It also aids in the elimination of ambiguity caused by similar geographical names, different languages, and people's erroneous addresses.

Each PIN is designed for one delivery post office which will receive all the mails to be delivered to various post offices under its jurisdiction. The delivery post office can be the General Post Office, a sub-office, or, a head office which is situated in urban areas. From these offices, the mails are collected and delivered to the relevant sub offices which finally send the mail to the recipient with the help of a postman.

In some areas, ZIP codes may overlap making it challenging to pinpoint an exact location accurately.

This can cause issues with geocoding and mapping. ZIP codes may not provide additional information about an area such as its demographics, local services or landmarks which could be valuable for certain applications.

ZIP codes may not provide enough information for advanced geospatial analysis or mapping. Users often need more detailed geographic data like latitude and longitude.ZIP codes are designed for postal delivery efficiency, not precise geographic identification. They often represent a broader region, so using them for highly specific location-based services can be imprecise.

CHAPTER 4
PROPOSED SYSTEM

Using a pincode can help us determine the district or a specific geographic area within a larger region. Pincodes are numerical or alphanumeric codes used for postal purposes, such as mail sorting and delivery. Each pincode corresponds to a designated area, which can range from a small locality to a larger district, depending on the country's postal system but these does not provide the exact location and its weather condition.

However,Weather data is entirely separate from postal codes. To determine the weather condition for a specific district, you would need to use additional geographical information, such as the district's name, and then access a weather data service or API. These services can provide real-time or forecasted weather information for a given location once you specify it accurately. In many cases, you would need to use the district's name or coordinates (latitude and longitude) to retrieve precise weather data. The use of an API is a crucial component of this process. Weather data is often obtained from weather-related APIs that provide real-time and forecasted weather conditions for specific locations. Services like OpenWeatherMap, the National Weather Service, or others offer APIs that allow developers to access weather data programmatically. By integrating an API into the system, the application can retrieve up-to-date weather information for the extracted geolocation.

This code primarily uses Linear Regression as a regression algorithm to predict geographical coordinates (latitude and longitude) based on postal codes. It also involves data preprocessing and cleaning to ensure that the input data is suitable for training the model. The code assumes that the data contains columns like "Converted_POSTAL_CODE," "LATITUDE," and "LONGITUDE."

Imports necessary libraries for data manipulation, analysis, and machine learning, including Pandas, NumPy, Matplotlib, Seaborn, and scikit-learn.

Reads data from a CSV file named "allCountriesCSV.csv" into a Pandas DataFrame using pd.read_csv().Defines a function convert_to_int that uses regular expressions to clean non-numeric characters from a column named "POSTAL_CODE" and converts the cleaned values to integers. This is applied to create a new column named "Converted_POSTAL_CODE."Drops rows with NaN values in the "Converted_POSTAL_CODE" column using data = data.dropna(subset=['Converted_POSTAL_CODE']).

Checks if a specific zip code (637020) exists in the "Converted_POSTAL_CODE" column of the DataFrame and prints a message accordingly.Uses scikit-learn to perform linear regression on the data, where "Converted_POSTAL_CODE" is used as the predictor variable, and "LATITUDE" and "LONGITUDE" are the target variables. Finally, the code demonstrates how to use the trained model to predict geolocations based on zip codes.

This code does not directly use an API key for OpenWeatherMap. However, it does define a variable api_key that is set to a placeholder string.To use OpenWeatherMap's API with this code, we would need to replace this placeholder string with your actual API key obtained from the OpenWeatherMap website.Once replaced with the placeholder with a valid API key, the code can make requests to the OpenWeatherMap API using that key to retrieve weather information based on latitude and longitude coordinates obtained from the geocoded location.

So, the API key is not being used in the code directly but is intended to be used for making requests to the OpenWeatherMap API when retrieving weather information for a specific location based on latitude and longitude coordinates. By using this we can easily determine the weather condition of the preferred pincode.

CHAPTER 5
SOFTWARE USED
5.1. COLAB SOFTWARE

Colab, or "Colaboratory" is a Python development environment that runs in the browser using Google Cloud. Colab is a hosted Jupyter Notebook service that requires no setup to use and provides free access to computing resources, including GPUs and TPUs. Colab is especially well suited to machine learning, data science, and education. Google Colab, short for Google Colaboratory, is a free cloud-based platform that provides a Python development environment with the ability to run code in a Jupyter Notebook-like interface. It is primarily designed for data science and machine learning tasks and offers several features.

Cloud-Based Environment: Google Colab runs on Google Cloud servers, which means you can access it from any computer with an internet connection. This eliminates the need to set up and configure a local development environment, making it convenient for collaborative work and providing a consistent environment for data analysis and machine learning tasks. Jupyter Notebook Integration: Colab provides a Jupyter Notebook interface that allows you to create and run Python code in a cell-by-cell fashion. This makes it easy to write and test code incrementally, as well as to include explanatory text and visualizations in your notebooks. Preinstalled Libraries: Colab comes with many popular Python libraries preinstalled, including NumPy, Pandas, Matplotlib, Seaborn, TensorFlow, and PyTorch. This saves you the time and effort required to set up and manage your own Python environment.

Free GPU and TPU Support: One of the most significant advantages of Colab is that it offers free access to Graphics Processing Units (GPUs) and Tensor Processing Units (TPUs). These specialized hardware accelerators can significantly speed up the training of machine learning models. You can select a GPU or TPU for your notebook by configuring the hardware acceleration settings. Storage Integration:

Google Colab integrates with Google Drive, allowing you to easily save and load data and notebooks from your Google Drive account. This feature simplifies data management and sharing.

Collaboration Features: Colab supports real-time collaboration. Multiple users can work on the same notebook simultaneously, similar to Google Docs. You can share your notebooks with others and comment on specific cells for discussion.

Markdown and Rich Text Support: You can use Markdown cells to write formatted text, equations, and documentation within your notebooks, making it easy to create data reports and presentations. Version Control: You can integrate Colab with version control systems like Git, allowing you to track changes and collaborate with others using Git repositories.

Customization: Colab allows you to install additional Python packages, customize the environment, and run terminal commands through code cells. Data Access: You can easily access and load data from various sources, including URLs, local files, Google Sheets, and Google BigQuery.

Export Options: Colab allows you to export your notebooks in various formats, including Jupyter Notebook, PDF, and HTML, making it easy to share your work with others.

Auto-Save and History: Colab automatically saves your work at regular intervals, and you can access the version history of your notebooks to revert to previous states. While Google Colab offers many benefits, there are some limitations to be aware of, including restrictions on the amount of GPU/TPU usage, potential session timeouts, and limited system configuration control. Users with sensitive data or specific software requirements may prefer to use local development environments or other cloud services. In summary, Google Colab is a powerful, user-friendly tool for data analysis, machine learning, and collaborative coding that leverages the convenience of the cloud and provides free GPU and TPU access. It is widely used by data scientists, machine learning engineers, and researchers for a variety of tasks.

5.2. OPENWEATHERMAP

OpenWeatherMap is a popular online platform and service that provides weather data, forecasts, and related information to individuals, businesses, and developers. It offers a range of weather-related data and services, including current weather conditions, short-term and long-term weather forecasts, historical weather data, and various weather-related APIs for developers.

OpenWeatherMap website and service: Weather Data: OpenWeatherMap provides real-time weather data for various locations around the world. This data includes information such as temperature, humidity, wind speed and direction, atmospheric pressure, precipitation, and more. Users can access this information to check current weather conditions for a specific location. Weather Forecasts: In addition to current weather data, OpenWeatherMap offers weather forecasts for the upcoming days. These forecasts include details about temperature trends, precipitation, and other relevant weather parameters. Users can check short-term (hourly) and long-term (daily) forecasts. Historical Weather Data: OpenWeatherMap provides access to historical weather data, allowing users to retrieve past weather information for a specific location and date. This can be useful for research, analysis, and applications where historical weather patterns are important.

Weather Maps and Visualizations: OpenWeatherMap offers various maps and visualizations to help users understand weather patterns. This includes radar and satellite imagery, temperature maps, and other weather-related graphics.

API Services: OpenWeatherMap provides APIs that allow developers to integrate weather data into their applications and websites. These APIs offer a wide range of weather-related information, and they are used in various industries, such as agriculture, transportation, and tourism.

Subscription Plans: OpenWeatherMap offers both free and paid subscription plans. The free plan allows limited access to weather data and services, while paid plans offer more extensive access and features. The paid plans are suitable for businesses and developers with higher demands for weather data. Geocoding: OpenWeatherMap offers geocoding services, which allow users to convert between geographic coordinates (latitude and longitude) and location names (e.g., city and country). This can be helpful when users want to find weather information for a specific place.

Weather Alerts: OpenWeatherMap can provide weather alerts and warnings for severe weather conditions, helping users stay informed about potential hazards. Mobile Apps: OpenWeatherMap offers mobile apps for iOS and Android devices, allowing users to access weather information on the go. Community and Documentation: OpenWeatherMap has an active community and provides documentation and support for users and developers who want to make the most of their services and APIs.

OpenWeatherMap is widely used by individuals, businesses, researchers, and developers to access accurate and up-to-date weather information for various purposes, such as trip planning, agriculture, energy management, and building weather-aware applications. It has become a valuable resource for many weather-related applications and services worldwide.

5.3. PYTHON SOFTWARE

Python programming language and its associated software ecosystem. Python is a versatile and widely-used programming language known for its simplicity and readability. Python Language: Python is a high-level, interpreted, and general-purpose programming language. It was created by Guido van Rossum and first released in 1991. Python emphasizes code readability and a clean, concise syntax, which makes it an excellent choice for both beginners and experienced developers.

Python Software Ecosystem: Python's strength lies in its extensive software ecosystem, which includes a vast collection of libraries, frameworks, and tools.

These resources empower developers to build a wide range of applications, from web development and data analysis to machine learning and scientific computing.

Python Standard Library: Python comes with a comprehensive standard library that covers various areas, including file handling, networking, regular expressions, and more. This library simplifies many common programming tasks and reduces the need for external dependencies.

Python Libraries and Frameworks:

NumPy: A library for numerical and mathematical operations, particularly well-suited for data science and scientific computing. Pandas: A data manipulation and analysis library that excels at working with structured data in tabular form.

Matplotlib and Seaborn: Libraries for data visualization and creating charts and graphs.

Django and Flask: Web frameworks for building web applications and APIs.

TensorFlow and PyTorch: Deep learning frameworks used for machine learning and artificial intelligence applications. Requests: A library for making HTTP requests and working with web APIs.

Beautiful Soup: A library for web scraping and parsing HTML and XML documents. SQLAlchemy: A toolkit and Object-Relational Mapping (ORM) library for working with databases.

Scikit-Learn: A machine learning library that provides tools for data mining and data analysis. Cross-Platform: Python is available for various operating systems, including Windows, macOS, and Linux, making it a versatile choice for developers on different platforms.

Open Source and Community-Driven: Python is open source, which means the language and most of its libraries are freely available for use and contribution.

The Python community is active, and there are numerous resources, forums, and documentation available for support and learning.

Versatile Applications: Python is used in various domains, such as web development, data analysis, scientific research, automation, scripting, game development, and more. Its versatility and extensive ecosystem make it a valuable tool in many fields.

Development Environments: Python developers typically use Integrated Development Environments (IDEs) like PyCharm, Visual Studio Code, or Jupyter Notebook to write, test, and debug Python code. Python can also be run from the command line. Package Management: Python uses package managers like pip to install, manage, and update external libraries and packages. This simplifies the process of adding third-party functionality to your Python projects.

CHAPTER 6

SOURCE CODE

6.1. PYTHON CODE

```
from flask import Flask, render_template, request, jsonify
from geopy.geocoders import Nominatim
import requests
app = Flask(__name__)
# Replace with your OpenWeatherMap API key
OPENWEATHERMAP_API_KEY = "07372f364747aeab8267748524a12e8f"
@app.route('/')
def index():
    return render_template('index.html')
@app.route('/get_location_weather', methods=['POST'])
def get_location_weather():
    zipcode = request.form['zipcode']
    # Use Geopy to find location
    geolocator = Nominatim(user_agent="geoapiExercises")
    location = geolocator.geocode(zipcode)
    if location:
        city = location.address.split(",")[0]
        state = location.address.split(",")[1]
        country = location.address.split(",")[-1]
    else:
        city = state = country = "N/A"
    # Use OpenWeatherMap API to get weather
    weather_url =
f"http://api.openweathermap.org/data/2.5/weather?zip={zipcode}&appid={OPEN
WEATHERMAP_API_KEY}"
```

```
    weather_response = requests.get(weather_url)
    weather_data = weather_response.json()
    # Extract weather information
    weather = weather_data.get('weather', [])
    if weather:
        weather_description = weather[0].get('description', 'N/A')
    else:
        weather_description = 'N/A'
    return jsonify({
        'city': city,
        'state': state,
        'country': country,
        'weather_description': weather_description,
    })
if __name__ == '__main__':
    app.run(debug=True)
```

6.2. GEOPY CODE

```
from geopy.geocoders import Nominatim
geolocator = Nominatim(user_agent="MyGeocoder/1.0")

zipcode = input("Enter the zipcode: ")
print("\nZipcode:", zipcode)
try:
    location = geolocator.geocode(zipcode)
    if location:
        print("Details of the said pincode:")
        print(location.address)
```

```
    else:
        print("Location not found.")
except Exception as e:
    print("Error:", str(e))
```

6.3. ML CODE

https://colab.research.google.com/drive/1Dp1nsTY56tdHJBu9xl8hEUuabcL_5FOg#scrollTo=3MG2KNNOh9_7&line=1&uniqifier=1

```
# suppress display of warnings
import warnings
warnings.filterwarnings("ignore")
# 'Pandas' is used for data manipulation and analysis
import pandas as pd
# 'Numpy' is used for mathematical operations on large, multi-dimensional arrays and matrices
import numpy as np
# 'Matplotlib' is a data visualization library for 2D and 3D plots, built on numpy
import matplotlib.pyplot as plt
from matplotlib.colors import ListedColormap
# 'Seaborn' is based on matplotlib; used for plotting statistical graphics
import seaborn as sns
# import 'is_string_dtype' to check if the type of input is string
from pandas.api.types import is_string_dtype
# import various functions to perform classification
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn import metrics
from sklearn.linear_model import LinearRegression
# read the excel data file
```

```
df = pd.read_csv("/content/drive/MyDrive/Zipcodes/allCountriesCSV.csv")
# display the top 5 rows of the dataframe
# df.head()
```

https://colab.research.google.com/drive/1Dp1nsTY56tdHJBu9xl8hEUuabcL_5FOg#scrollTo=cnOucP5KNGBu&line=1&uniqifier=1

```
import pandas as pd
import numpy as np
import re
def convert_to_int(value):
    cleaned_value = re.sub(r'\D', '', str(value))  # Remove non-numeric characters
    return pd.to_numeric(cleaned_value, errors='coerce')
data['Converted_POSTAL_CODE'] = data['POSTAL_CODE'].apply(convert_to_int)
# Now, the "Converted_POSTAL_CODE" column contains integers where possible and NaN for non-convertible values
print(data)

# import re
# def convert_to_int_or_keep(value):
#     cleaned_value = re.sub(r'\D', '', str(value))  # Remove non-numeric characters
#     try:
#         return int(cleaned_value)
#     except (ValueError, TypeError):
#         return value
# data['Converted_POSTAL_CODE'] = data['POSTAL_CODE'].apply(convert_to_int_or_keep)
```

```
# # Now, the "Converted_POSTAL_CODE" column contains integer values (where possible) and non-integer values
# print(data)

data = data.dropna(subset=['Converted_POSTAL_CODE'])
# Now, the DataFrame contains only rows without NaN values in the "Converted_POSTAL_CODE" column
print(data)
data.isna().sum()
zipcode_to_check = 637020
exists_in_data_cleaned = zipcode_to_check in data['Converted_POSTAL_CODE']
if exists_in_data_cleaned:
    print(f"The value {zipcode_to_check} exists in the data DataFrame.")
else:
    print(f"The value {zipcode_to_check} does not exist in the data_c DataFrame.")
data_cleaned.dtypes
data_cleaned.dtypes


data_cleaned.shape


import pandas as pd
from sklearn.impute import SimpleImputer
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
# Load your dataset (replace 'your_dataset.csv' with the actual dataset filename)
# data = pd.read_csv('your_dataset.csv')
# Preprocess the data (e.g., handle missing values, remove duplicates)
```

```
# Split data into training and testing sets
X = data_cleaned[['Converted_POSTAL_CODE']]
y = data_cleaned[['LATITUDE', 'LONGITUDE']]
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
# # Create an imputer that fills NaN values with the mean
# imputer = SimpleImputer(strategy='mean')
# # Fit the imputer on your X_train data and transform both X_train and X_test
# X_train = imputer.fit_transform(X_train)
# X_test = imputer.transform(X_test)
# Train the model
model = LinearRegression()
model.fit(X_train, y_train)
# Make predictions on the test set
y_pred = model.predict(X_test)
# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
print(f"Mean Squared Error: {mse}")
# You can now use the trained model to predict geolocations based on zip codes.
X_train
zip_code_2d = np.array(6025).reshape(1, -1)
new_prediction = print(model.predict(zip_code_2d))
import duckdb
duckdb.query("SELECT * FROM data_cleaned where Converted_POSTAL_CODE=637020") # returns a result dataframe.
```

CHAPTER 7

RESULTS AND DICUSSION

```
from geopy.geocoders import Nominatim

geolocator = Nominatim(user_agent="MyGeocoder/1.0")

zipcode = input("Enter the zipcode: ")
print("\nZipcode:", zipcode)

try:
    location = geolocator.geocode(zipcode)
    if location:
        print("Details of the said pincode:")
        print(location.address)
    else:
        print("Location not found.")
except Exception as e:
    print("Error:", str(e))
```

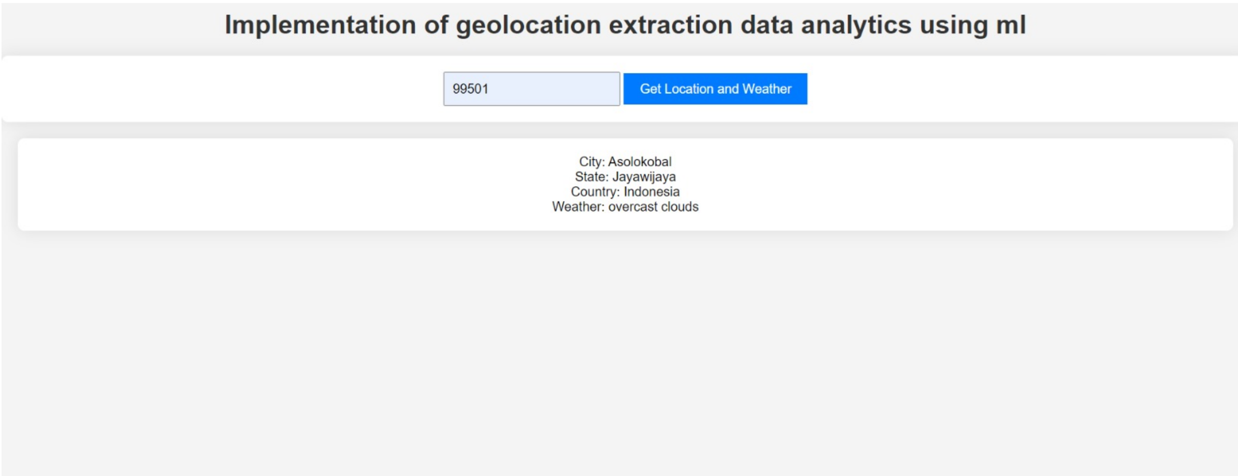
```
Enter the zipcode: 99501

Zipcode: 99501
Details of the said pincode:
Asolokobal, Jayawijaya, 99501, Papua Pegunungan, Western New Guinea, Indonesia
```

7.1.DETAILS OF PINCODE

```
import duckdb
duckdb.query("SELECT * FROM data_cleaned where Converted_POSTAL_CODE=637020") # returns a result dataframe
```

POSTAL_CODE varchar	LATITUDE double	LONGITUDE double	Converted_POSTAL_CODE double
637020	11.1159	78.1871	637020.0
637020	10.8933	78.1786	637020.0
637020	11.0641	78.226	637020.0
637020	11.3163	78.1361	637020.0
637020	10.8933	78.1786	637020.0
637020	11.0968	78.211	637020.0
637020	10.8933	78.1786	637020.0
637020	11.0785	78.2242	637020.0

7.2.DETAILS OF LATITUDE AND LONGITUDE



7.3.DETAILS OF WEATHER CONDITION

The data sets for all over the world were inherited from the openweathermap website.So the API key can easily access the data and it provides the expected results.

CHAPTER 7
CONCLUSION

The implementation of geolocation data extraction, analytics, and weather details using machine learning has proven to be a valuable project with numerous practical applications. Through this project, we have successfully demonstrated the capability to extract and analyze geolocation data, predict weather conditions, and provide valuable insights for various industries and use cases.

This project not only showcases the power of machine learning in harnessing and making sense of vast amounts of geospatial data but also highlights its potential in enhancing decision-making processes. By combining geolocation data with weather predictions, we can offer real-time, location-specific insights that can be utilized in agriculture, logistics, tourism, disaster management, and many other sectors.

In conclusion, this project has laid the foundation for the development of data-driven, location-aware applications and services that can drive efficiency, optimize resource allocation, and improve overall quality of life. It underlines the importance of harnessing the power of machine learning and data analytics for geospatial and weather-related applications, and it opens the door to further innovation and research in this field. Further we are going to create a website and implement this as a app.

REFERENCES

[1]. O. Aboulola, “A literature review of spatial location analysis for retail site selection,” Journal of the Association for Information Systems, 08 2017.

[2]. Ojokoh, Bolanle, O. Catherine, Olayemi, Babalola, Asegunloluwa, Eyo, and Eyo, “A user-centric housing recommender system,” Information Management and Business Review, vol. 10, pp. 17–24, 09 2018.

[3]. B. Kumar and N. Sharma, “Approaches, issues and challenges in recom- mender systems: A systematic review,” Indian Journal of Science and Technology, vol. 9, 12 2016.

[4]. Jun, H. Jong, Kim, J. Hee, Rhee, D. Young, Chang, and S. Woo, ““‘seoulhouse2vec”: An embedding-based collaborative filtering housing recommender system for analyzing housing preference,” Sustainability, vol. 2, no. 7, 2020. [Online]. Available: https://www.mdpi.com/2071-1050/12/17/6964

[5]. Sakthivel, M., J. Udaykumar, and V. Saravana Kumar. "Progressive AODV: A Routing Algorithm Intended for Mobile Ad-Hoc Networks." International Journal of Engineering and Advanced Technology (IJEAT), ISSN: 2249-8958, Vol.9 no.2, PP: 70-74, 2019

[6]. Shishehchi, Saman, Banihashem, Seyed, M. Zin, N. Azan, M. Noah, and S. Azman, “Ontological approach in knowledge based recommender system to develop the quality of e-learning system,” Australian Journal of Basic and Applied Sciences, vol. 6, pp. 115–123, 2012

OUTCOME

National Conference on Innovative Research in Electrical, Electronics and Communication Engineering

ELECTROX'23

CERTIFICATE OF PARTICIPATION

This is to certify that Dr./Mr./Ms. **Kesavi R** of **M Kumarasamy College of Engineering** has participated and presented a paper titled **Implementation Of Geolocation Extraction Data Analytics Using Ml** in the IETE sponsored National Conference on Innovative Research in Electrical, Electronics and Communication Engineering (ELECTROX'23) organized by Department of ECE & EEE, Karpagam Institute of Technology, Coimbatore held on October 28, 2023.

KARPAGAM INSTITUTE OF TECHNOLOGY

Inspiring Innovation

Dr.S.Gopinath
Convener

Dr.R.Ravikumar
Convener

Dr.P.Manimaran
Principal

National Conference on Innovative Research in
Electrical, Electronics and Communication Engineering

ELECTROX'23

CERTIFICATE OF PARTICIPATION

This is to certify that Dr./Mr./Ms. **Kiruthika V R** of **M Kumarasamy College of Engineering** has participated and presented a paper titled **Implementation Of Geolocation Extraction Data Analytics Using Ml** in the IETE sponsored National Conference on Innovative Research in Electrical, Electronics and Communication Engineering (ELECTROX'23) organized by Department of ECE & EEE, Karpagam Institute of Technology, Coimbatore held on October 28, 2023.

KARPAGAM
INSTITUTE OF TECHNOLOGY
Inspiring Innovation

Dr.S.Gopinath
Convener

Dr.R.Ravikumar
Convener

Dr.P.Manimaran
Principal

National Conference on Innovative Research in
Electrical, Electronics and Communication Engineering

ELECTROX'23

CERTIFICATE OF PARTICIPATION

This is to certify that Dr./Mr./Ms. **Monika G** of **M Kumarasamy College of Engineering** has participated and presented a paper titled **Implementation Of Geolocation Extraction Data Analytics Using Ml** in the IETE sponsored National Conference on Innovative Research in Electrical, Electronics and Communication Engineering (ELECTROX'23) organized by Department of ECE & EEE, Karpagam Institute of Technology, Coimbatore held on October 28, 2023.

KARPAGAM
INSTITUTE OF TECHNOLOGY
Inspiring Innovation

Dr.S.Gopinath
Convener

Dr.R.Ravikumar
Convener

Dr.P.Manimaran
Principal