A REPORT

ON

"GEOGRAPHICAL RISK PREDICTION MODEL"

BY

AKSHAY KUMAR JAIN **210C2030006**

Prepared in the partial fulfilment of the

Practice School II Course

AT

MOVIDU TECHNOLOGY PRIVATE LIMITED
BENGALURU, KARNATAKA

A Practice School II Station of

BML MUNJAL UNIVERSITY

BML MUNJAL UNIVERSITY
PRACTICE SCHOOL – II

JOINING REPORT

Date: 12/06/2023

Name of the Student	Akshay Kumar Jain
Name and Address of the Practice School – II Station	Name:- Inmovidu Tech Address:- 22nd Cross Rd, Sector 3, HSR Layout, Bengaluru, Karnataka 560102
Date of Joining PS-II station as per offer letter	29/05/2023
Actual date of reporting to PS-II station	03/07/2023
Department Allocated	Data Analysis
Name and Designation of the Industry Guide/ Industry Mentor for the Project	Ms. Mridula Chhetri Co- Founder
Industry Mentor Contact No. (If available)	9902803001
Industry Mentor E-mail Address (Compulsory)	mridula@inmovidutech.com

ACKNOWLEDGEMENT

The successful culmination of this endeavour has been achievable through the immeasurable backing and guidance rendered by numerous individuals. I sincerely acknowledge their notable contributions with profound gratitude. My heartfelt thanks extend to the Head of the organization, Vice Chancellor, Prof. Shyam Menon, and the Dean of SOET, Prof. (Dr) Maneek Kumar, for fostering an enabling atmosphere and essential resources that facilitated the seamless execution of PS-2 and this undertaking. I am equally beholden to the PS program Coordinators at the institution, namely Dr. Alipta Dutta, Dr. Kiran Khatter, Dr. Yogesh Gupta, and Dr. Ranbir Singh, for their unwavering support and unwavering guidance throughout the entire project journey. Special recognition is due to Mrs. Mridula Chhetri, Mr. Hareram Jha, and Mr. Amir Iqbal, the project's proficient experts, whose invaluable insights and proficiency significantly influenced the project's trajectory. I am appreciative of my University Faculty Mentor, Dr. Kiran Sharma, for providing perceptive feedback, motivation, and resolute support that profoundly contributed to the project's triumph. Furthermore, I extend my sincere gratitude to my fellow colleagues for their synergistic ethos and collaborative effort, which substantially enhanced the project's outcomes.

AKSHAY KUMAR JAIN

ABSTRACT

The Geographical Risk Prediction Model (GRPM) aims to revolutionize disaster preparedness and risk management strategies in India. By utilizing extensive data analysis and predictive modeling, the project seeks to develop a reliable and accurate risk prediction model for India and its states and union territories. Through a multi-faceted approach, the model extracts data from the different csv files made for weather and news sources and uses powerful tools like Tableau and Python libraries for analysis. The Geographical Risk Prediction Model's user interface, promises an intuitive and user-friendly experience for authorities.

Challenges related to data collection and processing have been addressed, ensuring the model's reliability. Expected outcomes include accurate predictions of natural disasters with binary indications. Leveraging available resources through government websites and news, the project was targeted for completion within two months.

The Geographical Risk Prediction Model holds the potential to safeguard lives and resources, contributing significantly to disaster management and risk reduction in India. This concise abstract effectively encapsulates the core essence of the project, outlining its methodologies, expected outcomes, and significance. It serves as an informative guide, providing a comprehensive overview of the report's key aspects

and highlighting the model's potential impact on disaster management and community resilience.

I was working under my industry mentor Ms. Mridula Chhetri who is the co-founder of the company.

OBJECTIVES

The objective of the Geographical Risk Prediction Model (GRPM) project is to develop a reliable and accurate risk prediction model for India and its states and union territories. The model will use extensive data analysis and predictive modelling to identify areas that are at high risk of natural disasters. This information will be used to improve disaster preparedness and risk management strategies in India.

Specifically, the GRPM project will:

- Extract data from weather and news sources to identify potential risk factors for natural disasters.
- Use powerful tools like Tableau and Python libraries to analyse the data and develop a risk prediction model.
- Create a user-friendly interface for the model that allows authorities to easily visualize and understand the risk predictions.
- Address challenges related to data collection and processing to ensure the reliability of the model.
- Produce accurate predictions of natural disasters with binary indications.
- Complete the project within two months.

The GRPM project has the potential to revolutionize disaster preparedness and risk management strategies in India. By providing accurate predictions of natural disasters, the model can help authorities to take steps to mitigate the risks and protect lives and property. The GRPM project also has the potential to contribute significantly to disaster risk reduction in India. By identifying areas that are at high risk of natural disasters, the model can help authorities to focus their resources on these areas and reduce the impact of future disasters.

Table of Contents

1. Brief Introduction of Organization's Business Sector

Education sector: -The college education sector in India encompasses a diverse range of undergraduate and postgraduate programs in various disciplines, offered by government-funded and private colleges as well as autonomous institutions. Students can choose from a wide array of courses, including arts, science, commerce, engineering, medicine, law, and social sciences.

The quality of education, faculty expertise, and infrastructure can vary across colleges, with prestigious institutions known for academic excellence and research contributions. College life offers a blend of academic learning and extracurricular activities, and internships and placements are emphasized to enhance students' employability. Challenges include maintaining consistent quality, improving research output, and addressing affordability concerns for economically disadvantaged students.

2. Brief History

Founding (2021): Movidu Tech Bengaluru was founded on October 4, 2021, by Mridula Chhetri and Pushp Raj Singh. The company is headquartered in Bengaluru, Karnataka.

Early years (2021-2022): In the early years, Movidu Tech Bengaluru focused on providing Microsoft licenses to educational institutions across India. The company also started offering internship programs to students, giving them the opportunity to gain real-world experience in the tech industry.

Recent years (2022-present): In recent years, Movidu Tech Bengaluru has expanded its services to include talent matchmaking and project

management. The company now helps companies find, hire, and engage talent in their pipeline while they are still studying. Movidu Tech Bengaluru also helps companies build projects at a fraction of the cost by leveraging the skills and expertise of student interns.

As of 2023, Movidu Tech Bengaluru is a leading provider of tech solutions for educational institutions and businesses in India. The company has a team of experienced professionals who are committed to helping their clients succeed.

3. Business Size

The company has over 100 employees. InMovidu does not use any commodities in its operations. In terms of the number of stocks, InMovidu is not publicly traded, so it does not have any stocks. However, the company may has issued some shares to its founders and investors.

4. Product Lines

Microsoft Licensing: Movidu Tech Bengaluru provides Microsoft licenses to educational institutions and businesses across India. The company offers a variety of licenses, including Windows, Office, and Azure.

Student Internships: Movidu Tech Bengaluru offers internship programs to students from all over India. The company's internship programs give students the opportunity to gain real-world experience in the tech industry.

Talent Matchmaking: Movidu Tech Bengaluru helps companies find, hire, and engage talent in their pipeline while they are still studying. The company's talent matchmaking services help companies save time and

money in the hiring process.

Project Management: Movidu Tech Bengaluru helps companies build projects at a fraction of the cost by leveraging the skills and expertise of student interns. The company's project management services help companies ensure that their projects are completed on time and within budget.

5. Competitors

Technosys: - It is an IT solutions provider that offers a wide range of services, including software development, system integration, and cloud computing. They have a strong track record of success and are well-respected in the industry.

Versatile:-Infotech is a software development company that offers a wide range of services, including mobile app development, web development, and cloud computing. They have a team of experienced developers who can create high-quality software solutions that meet the specific needs of their clients.

Internshala: - It is an internship and online training platform that connects students with internship opportunities and helps them develop their skills. It is one of the most popular internship platforms in India, with over 1 million registered students and over 100,000 internship opportunities.

6. Brief Description of all the departments

Web Development: This department is responsible for developing and maintaining Movidu's web products and services. They work on a wide range of projects, from developing new websites to maintaining existing ones.

Human Resources: This department is responsible for the recruitment, hiring, and onboarding of new employees. They also work to ensure that Movidu's employees are happy and productive.

Data/Business Analyst: This department is responsible for collecting, analyzing, and interpreting data to help Movidu make better decisions. They work on a variety of projects, from developing new data models to conducting market research.

Sales: This department is responsible for selling Movidu's products and services to customers. They work to identify potential customers, and to develop and execute sales strategies.

7. PLAN OF INTERNSHIP PROGRAM

Internship Project: Geographical Risk Prediction Model Development

Internship Department: Data/Business Analyst

During my internship at Movidu Tech Bengaluru, I had the opportunity to work on an exciting project focused on developing a Geographical Risk Prediction Model. The project aimed to analyze and predict potential risks associated with specific geographic locations, offering valuable insights to clients seeking to make informed decisions about talent engagement and project development.

Start and End Dates:

My internship project commenced on 29th May 2023 and successfully concluded on 29th July 2023. The duration of the project spanned 2 months.

Duties and Responsibilities:

As part of the Geographical Risk Prediction Model Development project, my key duties and responsibilities included:

Data Collection and Preparation: I collaborated with the team to gather relevant data from various sources, including demographic information, economic indicators, and historical risk data. The data was then processed and prepared for analysis.

Model Development: With guidance from experienced peers and mentors, I actively contributed to the development of the Geographical Risk Prediction Model. The model employed machine learning algorithms to identify

patterns and correlations in the data, enabling us to make accurate risk predictions.

Model Evaluation and Validation: I participated in evaluating and validating the model's performance by comparing its predictions with historical risk data. This iterative process allowed us to fine-tune the model for enhanced accuracy.

Data Visualization using Tableau: To aid in the interpretation and communication of results, I utilized Tableau as a powerful visualization software. I created interactive dashboards that showcased different trends and patterns within the data, providing insightful business intelligence for the team leader and facilitating easy understanding of key value drivers for further analysis.

Collaboration and Support:

Throughout the internship, I benefited from a supportive and collaborative work environment. Regular meetings with the team facilitated progress updates, problem-solving discussions, and knowledge sharing. Peer guidance and mentorship played a vital role in enhancing my understanding of the project's complexities and refining my technical skills.

Communication:

Effective communication was maintained through various channels, including WhatsApp messages and regular evening check-ins. These channels ensured seamless coordination, prompt responses to queries, and enhanced team cohesion.

Data Quality and Completeness:

To address challenges related to data quality and completeness during the model development process, the team divided into subgroups. We conducted a thorough manual check of the data, cross-referencing it with the source data to ensure accuracy. Additionally, we employed data validation features in Excel and implemented data comparison methods to identify and resolve any discrepancies. These rigorous measures contributed to the reliability of the model's predictions.

Projects:

During the course of my internship I also worked on Hotel Booking Analysis project which was suggested to me by my industry mentor and also created a short video course on covid data analysis.

8. Problem Statement

The Geographical Risk Prediction Model aims to analyze and predict potential risks associated with specific geographic locations. This model is designed to provide valuable insights to clients seeking to make informed decisions about talent engagement and project development. By identifying high-risk regions or areas vulnerable to certain factors, businesses can mitigate potential challenges and plan their strategies more effectively.

The project involves gathering and processing data from various sources, such as demographic information, economic indicators, and historical risk data. As an intern, my role was to assist in data collection, data preparation, and data analysis using tools like Excel and Python. I actively participated in the development of the risk prediction model, which employed machine learning algorithms to identify patterns and correlations within the data.

The model harnesses the power of IMD, rainfall data and the past landslide data to predict critical risks i.e., floods. By analyzing historical data and defined parameters, the model effectively predicts the likelihood of flooding based on rainfall patterns. Leveraging advanced machine learning algorithms, the model provides valuable insights for informed decision-making, allowing businesses, authorities, and clients to strategize effectively and mitigate potential risks. The model's interactive Tableau dashboards facilitate easy understanding and visualization of key insights, empowering different business owners and companies to plan and respond to geographical risks.

9. Introduction

Flood Prediction Model :-

Natural disasters are a major threat to life and property in India. In recent years, India has experienced several devastating natural disasters, including floods, cyclones, and earthquakes. These disasters have caused widespread damage and loss of life.

The GRPM project aims to revolutionize disaster preparedness and risk management strategies in India. By utilizing extensive data analysis and predictive modelling, the project seeks to develop a reliable and accurate risk prediction model for India and its states and union territories.

The GRPM project will use a multi-faceted approach to develop the risk prediction model. The model will extract data from a variety of sources, including weather data, news reports, and government data. The data will be analysed using powerful tools like Tableau, MySQL Workbench, and Python libraries. The model will also be designed to be user-friendly and intuitive, so that it can be easily used by authorities to make informed decisions about disaster preparedness and risk management.

The GRPM project has the potential to make a significant impact on disaster management in India. By providing accurate predictions of natural disasters, the model can help authorities to take steps to mitigate the risks and protect lives and property. The GRPM project also has the potential to contribute significantly to disaster risk reduction in India. By identifying areas that are at high risk of natural disasters, the model can help authorities to focus their resources on these areas and reduce the impact of future disasters.

10. Main Text

Project Methodology and Design

10.1 Assumptions Made

Assumptions were formulated to establish the relationships between rainfall patterns, IMD data, geographical attributes, and the likelihood of floods and landslides. These assumptions underpin the subsequent development and testing of the model. Additionally, we assumed that the impact of logistics connectivity and population would vary across districts. To categorize the data into tiers, districts with a significant impact of logistic company connectivity and higher population density were marked as Tier 1, while others were classified as Tier 2 and Tier 3, reflecting their relatively lower impact. The threshold points for each tier were based on the past pattern of the floods and the minimum and the maximum points were decided for each tier by seeing the overall trend and range of the rainfall that caused flood in the particular tiers.

Experimental Work and Data Collection

Data collection involved sourcing IMD and rainfall data from reliable news and government sources along with the most famous news resources to collect the required data. District-wise data categorization and tier-based grouping was as per team lead criteria that required the data to be at district level in order to cover some important routes for analysis, they formed the foundation for the analysis. The experimental work encompassed a variety of visualizations, relationships and other trends and chart for risk profiles and various impacts and other business problems for having a deep dive of the data set considering all the possible relations between the data columns to ensure better understanding and results.

10.2 Algorithm and Model Development

Algorithm Presentation

The Geographical Risk Prediction Model utilized the logistic regression algorithm, as it is best suited for categorical values. Logistic regression allowed us to predict the probability of floods and landslides based on the categorical outcomes, providing valuable insights for risk assessment.

10.3 Case Studies and Activities

Selection of Focused Locations

Specific areas of interest were chosen for the case studies, focusing on regions prone to floods and landslides. The identification of past affected locations allowed us to validate the model's performance. The main focus was kept on the landslide prone tier 1 districts due to high impact on the business market

Model Training and Validation

The cross-validation score and training score were used to evaluate logistic regression algorithm for supervised learning.

Cross-validation is a resampling technique used to know how well a model will take a broad view to new, unseen data. The primary goal of cross-validation is to estimate the model's performance on an independent dataset and detect potential issues like overfitting or underfitting.

The training score is the evaluation of the model's performance on the training dataset. It measures how well the model fits the training data, which is the data it has been explicitly exposed to during the learning process. It gives an indication of how well the model has memorized the training data and can be a useful metric to monitor during model training.

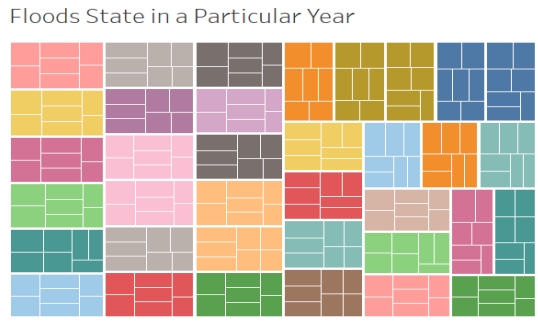
10.4 Results and Illustrations

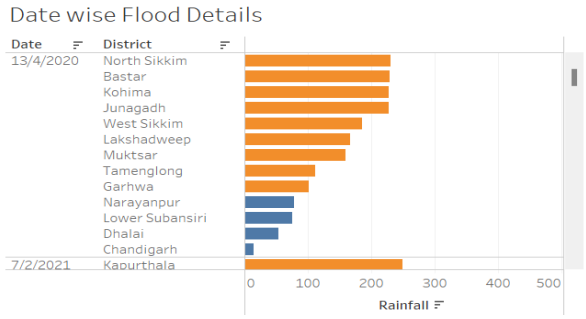
Performance Metrics

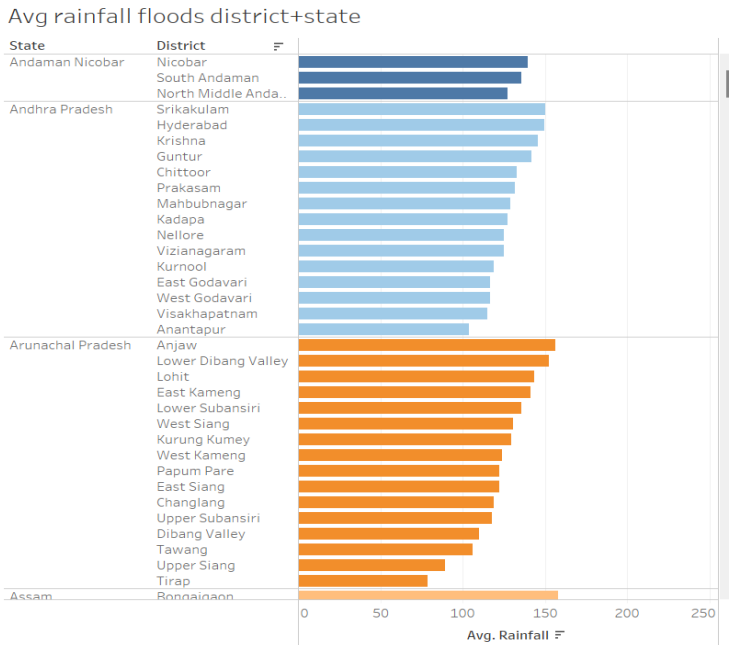
The Geographical Risk Prediction Model's performance was evaluated using various metrics, such as accuracy, cross validation score and training score. It depicted

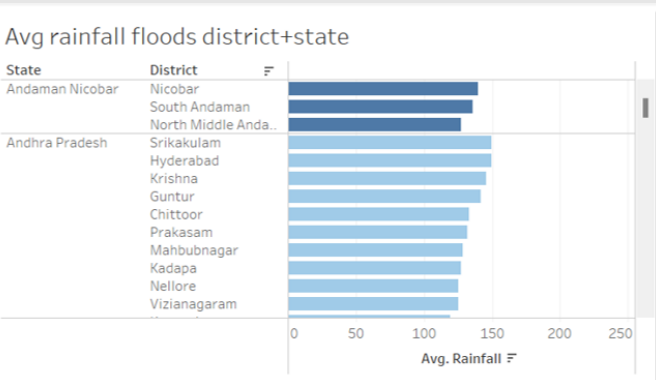
whether flood would occur or not based on districts, states and rainfall through the past data.

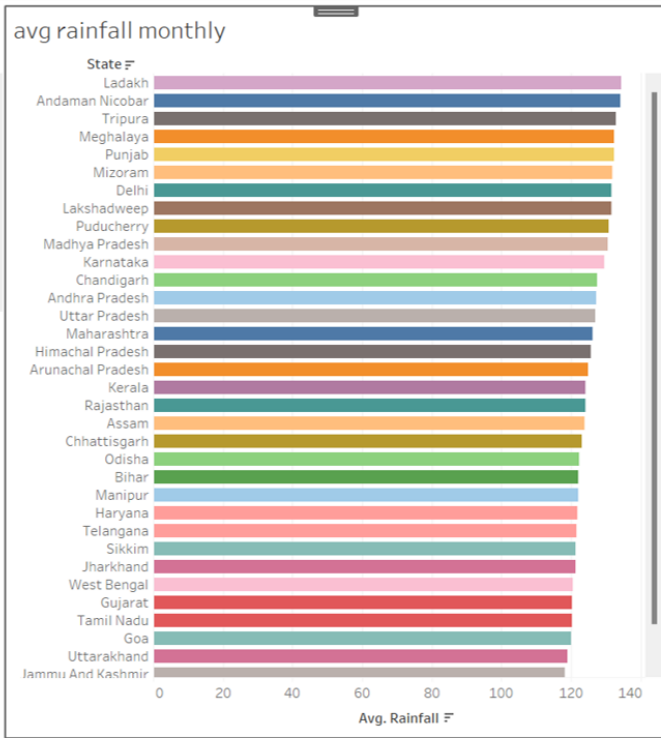
Visualizations: -



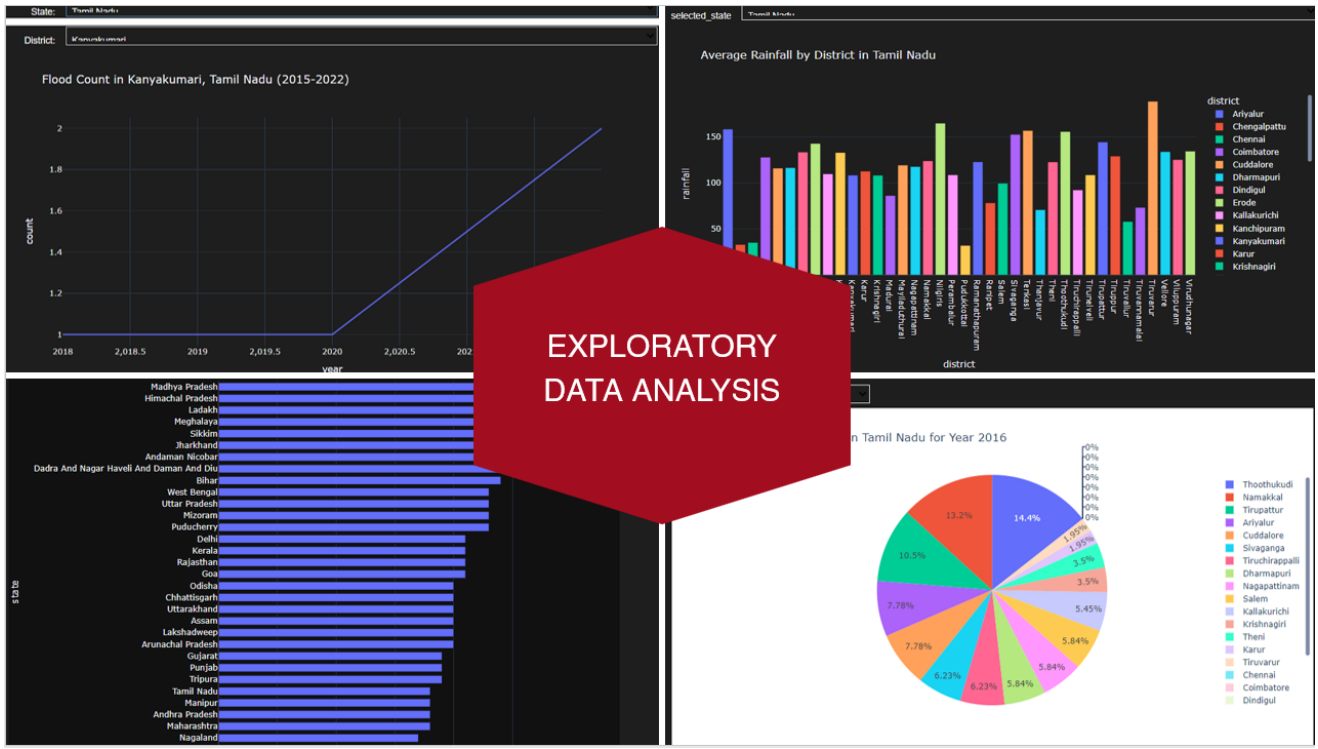









Avg Rainfall that can cause Flood by Month
month
Average Rainfall Year-wise
Year
EXPLORATORY
DATA ANALYSIS
Avg Rainfall(mm) in Chennai, Tamil Nadu (2015-2022)
rainfall
year
Tamil Nadu (2015-2022)
casualities
year


Flood Count in Kanyakumari, Tamil Nadu (2015-2022)
count
year
Average Rainfall by District in Tamil Nadu
rainfall
district
EXPLORATORY
DATA ANALYSIS
state
n Tamil Nadu for Year 2016

Visualization from the Data

Interactive Tableau dashboards were created to visualize how much rainfall in a particular year, Floods State in a Particular Year, count of Floods which occurred in a state and district in a particular year, Avg rainfall floods in district and state. illustrations were utilized to depict the key insights and critical information which could help the business take necessary actions and decisions.

10.5 Discussion and Interpretations

Discussion of Findings

The Geographical Risk Prediction Model exhibited significant predictive performance, accurately identifying regions with high flood and landslide risks. The analysis highlights several districts categorized as Tier 1, displaying a higher impact of logistic company connectivity and population density. These districts were consistently associated with elevated risks of floods and landslides, these outcomes helped to prevent clients take risky decisions.

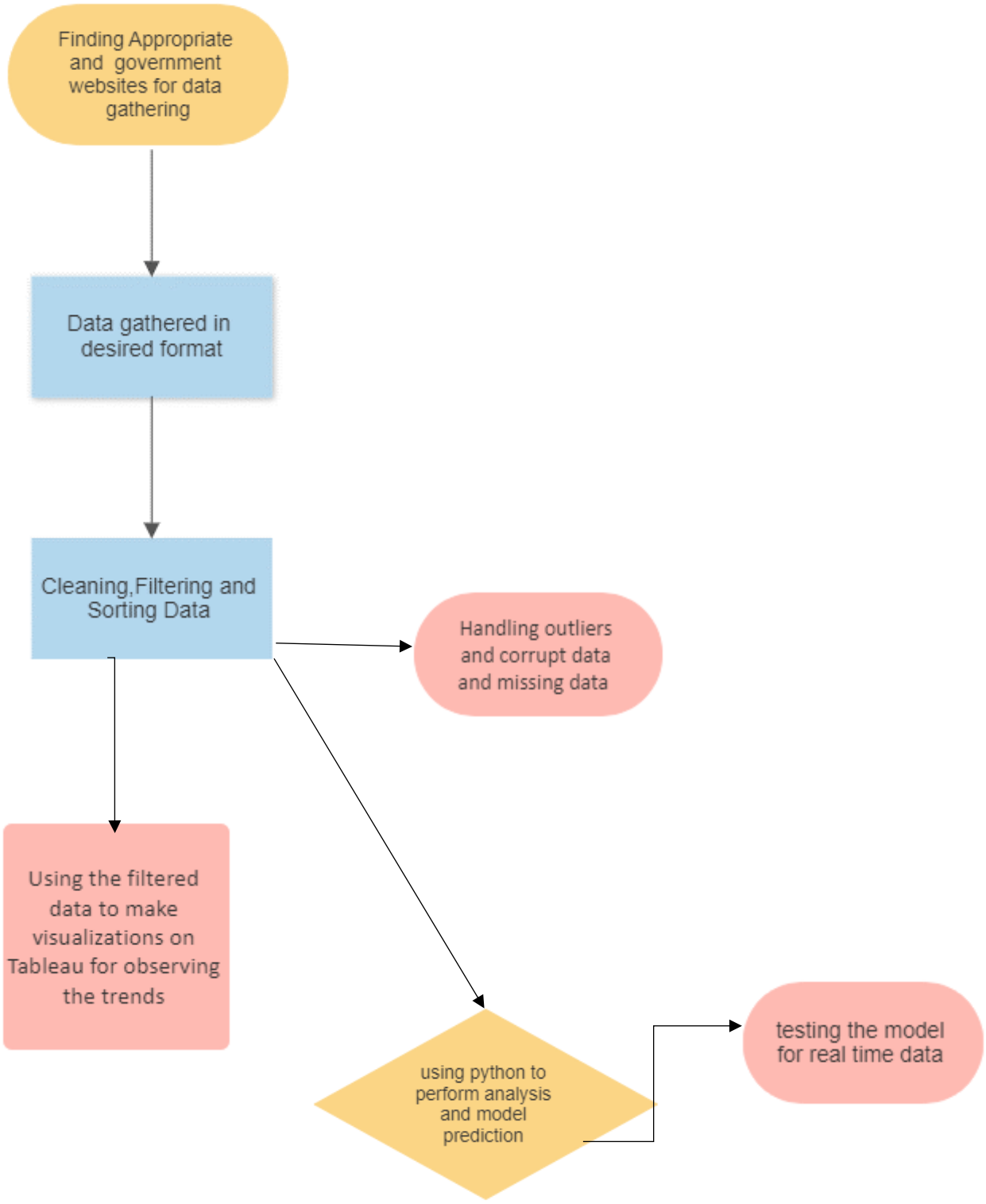
Furthermore, the model's predictions aligned with historical records of past affected locations, affirming its reliability and usefulness for disaster management and risk mitigation. The discussion dives into the specific factors contributing to the model's strong predictive capabilities, emphasizing the importance of district-wise data categorization in refining risk assessments.

Interpretation of Results

Interpreting the flood and landslide risk predictions, the model identified regions vulnerable to potential disasters. For districts marked as Tier 1, the probability of floods and landslides was notably higher, necessitating proactive measures for disaster preparedness. These results have implications for disaster management authorities, urging them to focus their resources and attention on high-risk regions to minimize potential damages.

Moreover, the interpretation considers the impact of logistic company connectivity and population density on risk assessments. By identifying correlations between these factors and flood/landslide probabilities, the model provides an understanding of risk patterns, supporting informed decision-making in resource allocation and continuous planning for future decisions and business insights.

10.6 Flow Chart of the Working



11.Outcomes of the Geographical Risk Prediction Model Project:

- Successful development of the Geographical Risk Prediction Model using logistic regression for accurate flood and landslide risk assessments.
- District-wise data categorization and tier-based grouping based on logistic company connectivity and population density proved effective in refining risk predictions.
- The model exhibited predictive analysis, identifying high-risk regions (Tier 1 districts) prone to floods and landslides.
- Consistent alignment of the model's predictions with historical records of past affected locations validated its reliability for disaster management.
- The Geographical Risk Prediction Model leveraged SQL for data extraction and manipulation, enabling the backend part of designing the website and fetching relevant information.
- Interactive Tableau dashboards provided visual representations of risk patterns, flood predictions, and landslide probabilities for easy comprehension.
- The Geographical Risk Prediction Model contributes valuable insights for informed decision-making and risk mitigation strategies.
- Potential implications of the model include improved disaster management practices and enhanced resilience in vulnerable regions.

12. Conclusion:

1. The Geographical Risk Prediction Model, leveraging logistic regression and tier-based categorization, effectively identified high-risk regions (Tier 1 districts) prone to floods and landslides. The model's alignment with historical records confirmed its reliability and suitability for disaster management.

2. District-wise data categorization played a pivotal role in refining risk predictions, allowing for targeted analyses and informed decision-making in disaster preparedness.

3. The integration of SQL facilitated data extraction and manipulation for the backend website, enabling users to access district-specific details, past landslide occurrences, and flood probabilities.

4. The model's insights provided valuable inputs for authorities to allocate resources efficiently, plan contingency strategies, and improve overall disaster management practices.

13.References

Government. (n.d.). *IMD*. Retrieved from Indian Meteorological Department: https://mausam.imd.gov.in/

MeT warns of heavy rain. (n.d.). Retrieved from Tribune India: https://www.tribuneindia.com/news/himachal/met-warns-of-heavy-rain-in-himachal-pradesh-till-saturday-529782

Rainfall Data. (n.d.). Retrieved from Data Government: https://data.gov.in/catalog/rainfall-india

Thrun, S. (n.d.). *Data Visualization with Tableau*. Retrieved from Udacity: https://learn.udacity.com/courses/ud1006

14. **PLAGIARISM REPORT**

ORIGINALITY REPORT

2%	1%	0%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Greenhouse Higher Secondary School Student Paper	1%
2	Submitted to Sultan Qaboos University Student Paper	1%
3	vocal.media Internet Source	<1%
4	Submitted to BML Munjal University Student Paper	<1%
5	www.reportsanddata.com Internet Source	<1%
6	rideally.com Internet Source	<1%

Exclude quotes	On	Exclude matches	< 6 words
Exclude bibliography	On		

PS2

ORIGINALITY REPORT

2%	1%	0%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Greenhouse Higher Secondary School Student Paper	1%
2	Submitted to Sultan Qaboos University Student Paper	1%
3	vocal.media Internet Source	<1%
4	Submitted to BML Munjal University Student Paper	<1%
5	www.reportsanddata.com Internet Source	<1%
6	rideally.com Internet Source	<1%

Exclude quotes	On	Exclude matches	< 6 words
Exclude bibliography	On		