Department of Computer Science & Engineering

National Institute of Technology, Srinagar

Hazratbal, Srinagar, Jammu & Kashmir-190006, India

PROJECT REPORT ON :

DESIGNING DISASTER-RESILIENT HOUSING IN INDIA:

A MACHINE LEARNING APPROACH FOR MATERIAL RECOMMENDATION, RISK PREDICTION, AND COST OPTIMIZATION

(ML INTERNSHIP)

SUBMITTED BY: NADIA KHAN

BRANCH: COMPUTER SCIENCE AND ENGINEERING (5TH SEM)

INSTITUTE: ISLAMIC UNIVERSITY OF SCIENCE AND TECHNOLOGY

PROJECT REPORT:

PROJECT TITLE:

Designing Disaster-Resilient Housing in India: A Machine Learning Approach for Material Recommendation, Risk Prediction, and Cost Optimization

TABLE OF CONTENTS:

➢ ABSTRACT

In India, natural disasters such as floods, earthquakes, and cyclones frequently devastate homes, displacing communities and causing widespread damage. A significant factor contributing to this vulnerability is the insufficient use of disaster-resistant materials and lack of resilient housing designs. This project seeks to develop an innovative, data-driven approach to recommend housing solutions that are both affordable and capable of withstanding such disasters. By analyzing local climate patterns, disaster risks, and material properties, machine learning models are used to identify optimal materials, predict disaster impacts, and optimize construction costs. The project combines data analytics and expert knowledge to create tailored recommendations for specific regions. Ultimately, the goal is to provide communities with practical, cost-effective housing designs that enhance resilience, reduce the socio-economic impacts of natural disasters, and contribute to sustainable development.

➢ PROBLEM STATEMENT

1. **Impact of Natural Disasters in India:**
   India is frequently impacted by natural disasters, such as floods, earthquakes, cyclones, and droughts, resulting in widespread destruction of homes and communities. Rebuilding these homes is not only costly but emotionally draining for those affected.

2. **Challenges with Existing Housing in Disaster-Prone Areas:**
   Many homes in these regions are built with inadequate materials that cannot withstand such disasters, leading to high repair and reconstruction costs. Additionally, current housing designs often fail to offer adequate protection, leaving homes vulnerable.

3. **High Cost and Unavailability of Disaster-Resilient Materials:**
   Disaster-resistant materials like reinforced concrete are often expensive or hard to source in some areas. This drives communities to rely on cheaper, less durable materials that increase the risk of damage during future disasters.

4. **Cycling of Vulnerability:**
   Families in disaster-prone areas often have to rebuild repeatedly, which keeps them stuck in a cycle of poverty and vulnerability. Short-term relief efforts are prioritized, but long-term prevention strategies are often overlooked.

5. **Need for Region-Specific Recommendations:**
   Housing designs must be tailored to specific regional risks, including local climate, hazards, and material availability. This project uses machine learning to provide customized material recommendations, predict disaster risks, and optimize affordable, disaster-resilient housing solutions.

➢ PROJECT GOAL

The project seeks to address these challenges by leveraging machine learning to:

- Recommend the most suitable construction materials based on disaster type, climate zone, and soil type.
- Predict disaster risk levels using environmental factors, helping communities understand their vulnerability.
- Optimize material costs to ensure affordable, durable, and disaster-resilient housing solutions.

➢ DATASET INFORMATION

I created this dataset based on real-world research focused on natural disasters, construction materials, and climate conditions. It includes data gathered from various sources, such as government disaster records, meteorological reports, and material cost studies. The dataset incorporates key features needed to design housing solutions that are not only disaster-resilient but also affordable and region-specific.

Description of Each Feature

1. **Disaster Type:** The type of disaster (e.g., flood, earthquake, cyclone), based on historical disaster data.
2. **Disaster Frequency:** How often disasters occur in a region, derived from disaster occurrence records.
3. **Risk Level:** The severity of the disaster in a region, based on historical impact data.
4. **Disaster Location:** Geographical regions most prone to specific types of disasters.
5. **Temperature (°C):** Average regional temperature, influencing material selection and design.
6. **Precipitation (mm):** Rainfall levels that impact housing safety and material durability.
7. **Wind Speed (km/h):** Wind conditions affecting structural stability, especially in cyclone-prone areas.
8. **Humidity (%):** Moisture levels that influence material longevity and resistance to mold.
9. **Climate Zone:** Classification of the region’s climate (e.g., tropical, arid), impacting design and material choices.
10. **Recommended Material:** The most suitable construction material for the region based on disaster risks and climate.
11. **Material Cost (per sqft):** The price of the recommended material, ensuring affordability.

12. **Material Durability:** Expected performance of materials under disaster conditions, sourced from material science studies.
13. **Soil Type:** Ground conditions affecting foundation stability, identified through local soil data.
14. **Structural Features:** Construction elements (e.g., reinforced walls, seismic bracing) that enhance resilience.
15. **Foundation Type:** Foundation structures best suited for different types of disasters.
16. **Income Level (annual):** Average income in the region, helping assess the affordability of materials.

How These Features Contribute to the Project

1. **Material Recommendation System**:
   - The features related to disaster type, climate conditions, soil type, and structural requirements are used to recommend the most suitable materials for a region. These recommendations aim to balance durability, cost, and the ability to withstand specific natural hazards.
2. **Risk Prediction**:
   - Features such as temperature, precipitation, wind speed, humidity, and disaster frequency are used to predict the level of disaster risk in different regions. This helps in identifying high-risk areas where stronger, more expensive building materials might be necessary.
3. **Cost Optimization**:
   - Material cost and income level are key for cost optimization. The model analyzes the cost of different materials and recommends affordable options that do not compromise the safety and durability of the structure. By considering income levels, the project ensures that the recommended housing solutions are financially feasible for the local population.

CODING:

➢ METHODOLOGY

❖ Data Collection:

- The dataset was created from research on natural disasters, climate conditions, materials, and housing design. The data underwent cleaning and preprocessing to ensure its suitability for machine learning analysis.

❖ Data Cleaning and Preprocessing:

1. **Handling Missing Values:**
   - The dataset is checked for missing values (NaN). Any rows containing missing values are identified and can be handled, in this case, by dropping them.
2. **Removing Duplicates:**
   - Duplicate rows, if present, are removed to ensure the dataset remains unique and does not affect the analysis or model performance.
3. **Date Conversion:**
   - The Disaster Date column is converted into a standardized datetime format. This is essential for ensuring consistency when handling dates in further analysis or modeling.
4. **Encoding Categorical Variables:**
   - Categorical columns such as Disaster Type, Risk Level, Disaster Location, etc., are encoded into numerical values using label encoding. This is necessary because machine learning models require numerical data to operate.
5. **Handling the 'Income Level (annual)' Column:**
   - Any commas in the Income Level (annual) column (representing thousands) are removed, and the column is then converted into a numeric format. Any values that cannot be converted are set to NaN.
6. **Standardizing Numerical Data:**
   - The Income Level (annual) column is standardized using a StandardScaler. This transformation ensures that the feature has a mean of 0 and a standard deviation of 1, making it easier for machine learning models to process.
7. **Normalizing Numerical Features:**
   - Several numerical columns, such as Disaster Frequency, Temperature (C), Wind Speed (km/h), etc., are normalized using a MinMaxScaler. This process scales all values to a range between 0 and 1, ensuring that no feature dominates due to its larger range.

❖ Exploratory Data Analysis (EDA):

1. **Count Plot: Disaster Type Frequency**
   - A count plot is used to visualize the distribution of disaster types in the dataset. It shows how frequently each disaster type occurs.
2. **Count Plot: Disaster Type vs. Risk Level**
   - A count plot is used to visualize the relationship between disaster types and their associated risk levels, highlighting how the risk varies across different disaster types.
3. **Count Plot: Climate Zone vs. Recommended Material**
   - This count plot helps visualize the distribution of climate zones, with the additional breakdown of recommended materials for each climate zone. It shows which materials are suitable for different climates.
4. **Bar Plot: Material Cost vs. Durability**
   - A bar plot is used to display the relationship between the average material cost and durability for different materials. It helps compare the cost-effectiveness and durability of materials.
5. **Pair Plot: Cost, Risk, and Disaster Frequency**
   - A pair plot is used to explore the relationships between material cost, disaster frequency, and risk level. It helps identify correlations between these features and understand how they interact with each other.
6. **Scatter Plot: Material Cost vs. Durability**
   - A scatter plot is used to visualize the relationship between the material cost and durability, with color representing the durability level. This helps in understanding how different materials compare in terms of cost and longevity.

➢ ALGORITHM USED

1. *Material Recommendation System*

Algorithm Used: Random Forest Classifier

- **Purpose**: To recommend suitable materials based on factors such as temperature, precipitation, wind speed, humidity, climate zone, disaster type, and risk level.

Steps:

1. **Features and Target Variables**:

- **Features**: Temperature, Precipitation, Wind Speed, Humidity, Climate Zone, Disaster Type, Risk Level.
- **Target**: Recommended Material (the material to be used for construction, e.g., wood, bamboo, concrete).

2. **Model**:
   - The RandomForestClassifier is used because it can handle categorical features well and can make accurate predictions for multi-class classification problems (e.g., predicting different types of materials).
   - It works by creating multiple decision trees and taking the majority vote for the final classification.
3. **Training and Testing**:
   - The dataset is split into training and testing sets (80% for training, 20% for testing).
   - The model is trained on the training data (material_model.fit()).
4. **Prediction**:
   - After training, the model is used to predict the material based on unseen data (material_model.predict()).

Output:

- **Predicted Material**: Based on the input features, the model outputs a material recommendation (e.g., reinforced concrete, wood, bamboo, etc.). This recommendation helps in designing the housing structure for the specific region and disaster type.

2. ***Risk Prediction***

Algorithm Used: Random Forest Classifier

- **Purpose**: To predict the disaster risk level based on environmental factors like temperature, precipitation, wind speed, humidity, and climate zone.

Steps:

1. **Features and Target Variables**:
   - **Features**: Temperature, Precipitation, Wind Speed, Humidity, Climate Zone.
   - **Target**: Risk Level (e.g., low, medium, high risk).
2. **Model**:
   - Similar to the material recommendation system, the RandomForestClassifier is used here as well to classify the risk level based on environmental and climatic data.
   - The algorithm builds multiple decision trees, each using a random subset of features, and aggregates their outputs for a more accurate classification.
3. **Training and Testing**:

- The dataset is split into training and testing sets. The model is trained on the training set (risk_model.fit()).

4. **Prediction**:
   - After training, the model predicts the risk level based on the environmental features.

Output:

- **Predicted Risk Level**: The model outputs the predicted disaster risk level (e.g., low, medium, high) for a specific region, which helps in determining how vulnerable a region is to certain types of disasters.

3. ***Cost Optimization (Material Cost Prediction)***

Algorithm Used: Linear Regression

- **Purpose**: To predict the material cost per square foot based on disaster-related factors such as temperature, precipitation, wind speed, humidity, climate zone, disaster type, and risk level.

Steps:

1. **Features and Target Variables**:
   - **Features**: Temperature, Precipitation, Wind Speed, Humidity, Climate Zone, Disaster Type, Risk Level.
   - **Target**: Material Cost (per sqft).
2. **Model**:
   - LinearRegression is used to predict continuous values, such as the material cost. Linear regression works by finding the relationship between the input features and the target variable through a linear equation.
   - The model is fitted to the training data to learn this relationship (cost_model.fit()).
3. **Training and Testing**:
   - Similar to the other tasks, the dataset is split into training and testing sets. The model is trained on the training data (cost_model.fit()).
4. **Prediction**:
   - The trained model predicts the material costs on unseen data (cost_model.predict()).

Output:

- **Predicted Material Cost**: The model predicts the cost of the materials (e.g., cost per square foot) based on the provided environmental factors. This helps in recommending the most cost-effective materials for disaster-resilient housing based on location and disaster risk.

These algorithms together work to **minimize disaster-related damages** by providing accurate, data-driven recommendations on materials, risk levels, and costs for disaster-resilient housing designs.

➢ LITERATURE REVIEW

The **Literature Review** for this project highlights key areas of existing research and points out the gap that this project aims to address. The main areas of focus include:

1. **Disaster Risk Assessment Models Using Machine Learning**
   - Previous research has explored the use of machine learning techniques for predicting and assessing the risk of natural disasters. These models use environmental data (e.g., temperature, precipitation, wind speed) to predict the likelihood and severity of disasters, helping communities better prepare.
2. **Studies on Cost-Effective, Disaster-Resistant Building Materials**
   - There are numerous studies that analyze and recommend building materials that can withstand natural disasters like floods, earthquakes, and cyclones. Research in this area emphasizes the need for durable materials that are also affordable and locally available.
3. **Application of Predictive Analytics in Climate and Construction Domains**
   - Predictive analytics has been applied in the fields of climate modeling and construction to forecast various factors that affect building safety. These studies use data on local climate, hazards, and materials to predict the best practices for disaster-resilient housing construction.

Gap in Existing Research:

While existing studies have concentrated on **disaster risk prediction**, they tend to focus on either risk assessment or material recommendations separately. There is a lack of integrated models that combine **cost optimization**, **material recommendations**, and **risk prediction** in a single framework for disaster-resilient housing. This project aims to fill this gap by providing a comprehensive, data-driven approach that integrates these three aspects to deliver more effective, affordable, and sustainable housing solutions.

➢ RESULTS AND ACCURACY

The model demonstrates high accuracy in predicting disaster risk levels, effectively classifying regions based on their vulnerability. It successfully recommends materials that align with expert guidelines, ensuring that the materials are suitable for withstanding specific disasters. Additionally, the cost estimation feature accurately identifies affordable materials without compromising their durability. The results show that the model can provide reliable, data-driven recommendations for designing safer and more resilient housing in disaster-prone areas.

➢ APPLICATIONS

1. **Governments**: Use the model to prioritize disaster-prone regions and plan housing projects based on risk levels.
2. **NGOs**: Leverage the material recommendations and cost estimations to implement affordable, disaster-resilient housing solutions.
3. **Housing Developers**: Design cost-effective homes using recommended materials that can withstand local disasters.
4. **Policy-Making**: Provide data-backed insights for creating building regulations and material standards that improve housing resilience.
5. **Disaster Risk Management**: Assist in identifying vulnerable areas and selecting appropriate construction materials for safer communities.

➢ FUTURE SCOPE

1. **Real-Time Data Integration**: Incorporating live climate and disaster data for dynamic recommendations.
2. **Advanced Machine Learning**: Implementing deep learning for improved prediction accuracy.
3. **Wider Regional Coverage**: Expanding to more regions and disaster types for broader applicability.
4. **Long-Term Resilience**: Analyzing material performance over time, considering aging and climate change.
5. **Cost-Effectiveness Over Time**: Including long-term maintenance costs for more comprehensive recommendations.
6. **Local Stakeholder Collaboration**: Refining recommendations based on feedback from local communities and experts.
7. **Policy Integration**: Integrating the model into national and international disaster management policies.

❖ *CROSS-DOMAIN APPLICATIONS*

1. **Disaster Management**: The model can be extended to broader disaster management systems, integrating with early warning systems and emergency response strategies for more proactive disaster prevention and mitigation.
2. **Urban Planning**: Urban planners can use the model to design cities that are resilient to natural disasters, integrating disaster-resistant housing with broader infrastructure and zoning policies.
3. **Environmental Science**: By incorporating climate change models, the system can help predict future disaster risks and their impact on housing, contributing to environmental sustainability.
4. **Supply Chain Management**: The recommendations for materials can be linked to supply chain systems to ensure the availability of affordable, disaster-resistant materials in vulnerable regions.
5. **Economic Modeling**: The cost optimization aspect can contribute to broader economic studies, particularly in assessing the economic impact of natural disasters and the long-term savings of investing in resilient infrastructure.
6. **Public Health**: Disaster-resilient housing can be aligned with public health strategies to reduce health risks in disaster-prone regions, ensuring safer living environments during and after disasters.

By bridging multiple domains, the model can provide a holistic approach to disaster resilience, benefiting not only housing but also related sectors like urban planning, public health, and economic development.

➢ CONCLUSION

This project uses machine learning to recommend disaster-resilient housing solutions, focusing on material selection, risk prediction, and cost optimization. By analyzing disaster types, climate conditions, and regional factors, the model offers practical, affordable housing recommendations. The project demonstrates potential for future improvements, such as real-time data integration and broader application across regions. With its cross-domain impact, it can contribute to disaster management, urban planning, and public health, providing a valuable tool for creating safer, more resilient communities.