**Ideation Phase**

**Defining the Problem Statements**

|  |  |
| --- | --- |
| **Date** | **26-09-2023** |
| **Project Name** | **Assessment of Marginal Workers in Tamil Nadu- A Socioeconomic Analysis** |

**Assessment of Marginal Workers in Tamil Nadu- A Socioeconomic Analysis**

**Problem Definition and Design Thinking**

**Introduction**

This analysis delves into the socioeconomic dynamics of marginal workers in Tamil Nadu, India. We will examine key demographics, including age, industrial category, and gender distribution, using insightful visualizations to highlight the challenges faced by this vulnerable group. By shedding light on their profiles, this study aims to inform targeted policies and initiatives that can uplift the marginalized workforce in Tamil Nadu.

**Problem Statement**

Objective: The aim of this data analytics project is to comprehensively assess and compare the living conditions and income levels of male and female marginal workers in the state of Tamil Nadu, India.

Data: We have a dataset containing various inputs about marginal workers(e.g., age, location, number of workers, education etc.) along with their corresponding incomes. This data will be used to train and evaluate our machine learning model.

**Key Challenges:**

1. Data Quality: Ensuring the dataset is clean, complete, and free of errors.

2. Feature Selection:Obtaining accurate and comprehensive data on marginal workers, especially disaggregated by gender.

3. Model Selection: Choosing the appropriate machine learning algorithm(s) for the task.

4. Security:Managing sensitive personal data while respecting privacy and ethical considerations is essential.

5. Deployment: Creating a user-friendly interface or API for end-users to make predictions.

**Design Thinking Approach**

**Empathize:**

To develop we need to empathize a comprehensive socio-economic analysis of marginal workers in Tamil Nadu, India, with a focus on understanding their demographic characteristics based on age, industrial category, and gender. Create informative visualizations to facilitate data-driven policy decisions and interventions aimed at improving the well-being of this vulnerable segment of the population.

**Actions:**

- Data from surveys conducted among marginal workers or their households to gather detailed information.

- Information on existing government programs and policies related to marginalized workers, as these may impact the data.

- Segment marginal workers based on age, gender, and industrial category to create profiles.

**Define:**

Based on our understanding of the problem and the users' needs, we will define clear objectives and success criteria for our project.

**Objectives:**

- Conduct Exploratory Data Analysis(EDA) to understand the distribution and characteristics of the data, identify outliers, and detect any missing values or inconsistencies.

- Develop predictive models to estimate income levels or employment status based on demographic and educational attributes, providing insights into the factors influencing these outcomes.

**Ideate:**

Brainstorm potential solutions and approaches to address the problem. This phase involves thinking creatively and considering various algorithms and techniques for socio-economic analysis.

**Actions:**

- Explore different machine learning algorithms such as decision trees, random forests.

-Experiment with feature engineering techniques to enhance model performance.

-Consider incorporating external data sources (e.g., educational attainment, skill levels, impact income and job opportunities) to improve analysis.

**Prototype:**

Create a prototype of the machine learning model and the user interface for socio-economic analysis.

**Actions:**

- Develop a IBM cognos analysis for data pre-processing, model training, and evaluation.

- Create a interface tools like Django to allow users to input details of marginal workers.

- Test the prototype with a subset of the dataset to ensure it meets performance objectives.

**Test:**

Evaluate the model's performance using appropriate metrics and gather feedback from users.

**Actions:**

- Split the dataset into training and testing sets.

- Train the model on the training set and evaluate it on the testing set.

­- We use metrics such as Age distribution, Gender-based, Industrial category, Educational based to maintain the performance of the model.

- Collect user feedback on the web interface for usability and accuracy.

**Implement**

When the prototype achieves its defined objectives and favourable feedback, proceed with the complete implementation.

**Actions:**

- Perform in-depth analysis and draw conclusions from the dataset.

- Train the final machine learning model using the entire dataset.

- Develop a production-ready web application that incorporates the model.

**Iterate**

Continues enhancement is necessary. Collect user feedback and iterate on both the model and interface to improve accuracy and user-friendliness in the context of assessing the socio-economic status of marginal workers in Tamil Nadu.

**Actions:**

- Continuously monitor the model's performance and periodically retrain it using the most recent data.

- Actively incorporate user feedback to enhance the web interface's usability and effectiveness.

- Continuously assess the long-term impact of your interventions on the socio-economic status of marginal workers. This involves tracking changes over time and adjusting strategies accordingly.

**Conclusion**

The assessment of marginal workers in Tamil Nadu is a crucial undertaking to address their socio-economic challenges. Through data analysis and a user-friendly web interface, this project aims to provide valuable insights and support evidence-based decision-making .Continuous improvement, user feedback, and model refinement are essential for the success and relevance of this initiative. Ultimately, this endeavor strives to improve the well-being and inclusion of marginal workers, fostering positive socio-economic changes in Tamil Nadu.