



COLLEGE CODE : 8107

COURSE : DATA ANALYTICS WITH COGNOS

PHASE II: PROJECT SUBMISSION

PROJECT TITLE: Assessment of marginal workers in
Tamil Nadu – A socioeconomic Analysis

TEAM MEMBERS DETAILS:

JAYASHREE. M
QUINY CHRISTABEL. R
SHINORA THOMAS. T
SHREENITHE. R

Introduction

In this innovation we transform the way data is extracted, cleaned, analyzed, visualized, and reported in the assessment of marginal workers' socio-economic conditions. It aims to make the entire process more efficient, insightful, and accessible.

Key Components:

1. Data Extraction Automation:

- Implementing machine learning algorithms to extract relevant socio-economic data from various sources, including surveys, databases, and interviews.
- In this project we are already given the data set.
- Using natural language processing (NLP) to extract textual data from documents, enabling faster and more accurate data collection.

2. Data Cleaning Assistance:

- We start this with data cleaning algorithms that can detect and rectify common data errors and inconsistencies automatically.
- Use of data validation rules and machine learning models (cross validation, outlier detection and handling, resampling) to suggest corrections, saving time and ensuring data accuracy.

3. Advanced Data Analytics:

- Integrate sophisticated machine learning model that can quickly identify patterns, correlations, and trends within the data.
- The machine learning models that can be used here are
 - Linear regression-predicting income based on education, age, employment status

- K-means clustering algorithm- helping to identify distinct segments within the population.

4. Interactive Data Visualization:

- Usage of data visualization techniques like heat maps, interactive charts, and geographic mapping to provide a holistic view of the socio-economic landscape.
- Through interpreting these techniques we are provided insights about data.

5. Automated Reporting Engine:

- Presenting the main finding of our analysis.
- Reports can be generated in various formats, including PDF, PowerPoint, or web-based dashboards.

Benefits:

- **Accuracy:**
Automation reduces the likelihood of human errors in data extraction and cleaning, ensuring high data accuracy.
- **Insights:**
Advanced analytics capabilities provide deeper insights into the socio-economic conditions of marginal workers, enabling evidence-based decision-making.
- **Interactivity:**
Interactive reports engage users and allow them to explore data dynamically, leading to deeper insights and more informed decision-making.
- **Enhanced Data Visualization:**
Innovative tools and techniques can help you create visually compelling and interactive data visualizations.