Industrial Human Resource Geo-Visualization

Approach:

- Merged all the csv data file provided to me and I created DataFrame.

- The classical machine learning tasks like Data Exploration, Data Cleaning, Feature Engineering, Model Building and Model Testing. Used Natural Language Processing for analyzing the various core industries and group the business categories like Retail, Poultry, Agriculture, Manufacturing etc.

Workflow :

The workflow of the project involves Human Resource Management and is conducted using Jupiter Notebook (VS Code).

- I merged all the CSV files to create a DataFrame, and then imported the necessary libraries into our Python environment.
- For data manipulation and analysis, I'm utilizing the Pandas library.
- import pandas as pd

Loaded the dataset:

- Next, I loaded the given dataset into a Pandas DataFrame, replacing "dataset.csv" with the filepath of the dataset (IHR).

- df = pd.read_csv('IHR.csv’)

- To gain insight into the dataset's structure and information, I viewed the first few rows of the dataset (IHR). By viewing the first few rows, I can get a sense of the dataset’s structure and the information it contains.
  print(df.head())

Loaded the dataset : -

- Summary statistics provide valuable insights into the distribution, central tendency, and variability of our data. print(df.describe()) Checking for missing values in the dataset (IHR) is crucial as they can affect the accuracy and reliability of our analysis. print(df.isnull().sum()) .

- Understanding the data types of each column is essential for data manipulation and analysis. Let’s examine the data types of each column. print(df.dtypes).

- "Data cleaning was performed using a lambda function to remove unwanted characters from the given dataset (IHR).

Feature Engineering :

- After completing the data cleaning process, the cleaned data was saved to a file named 'IHR.csv'. Then, feature engineering was performed using get_dummies().
- Additionally, line plots, scatter plots, and box plots were executed using Matplotlib and Seaborn.
- Following this, the sklearn library was utilized for model building and model testing.
- I imported a dataset of titles and descriptions from the dataset (IHR). Using seven columns, I tokenized them using NLTK and Python.

Natural Language Processing :

- It is the branch of Artificial Intelligence that gives the ability to machine understand and process human languages. Human languages can be in the form of text or audio format.

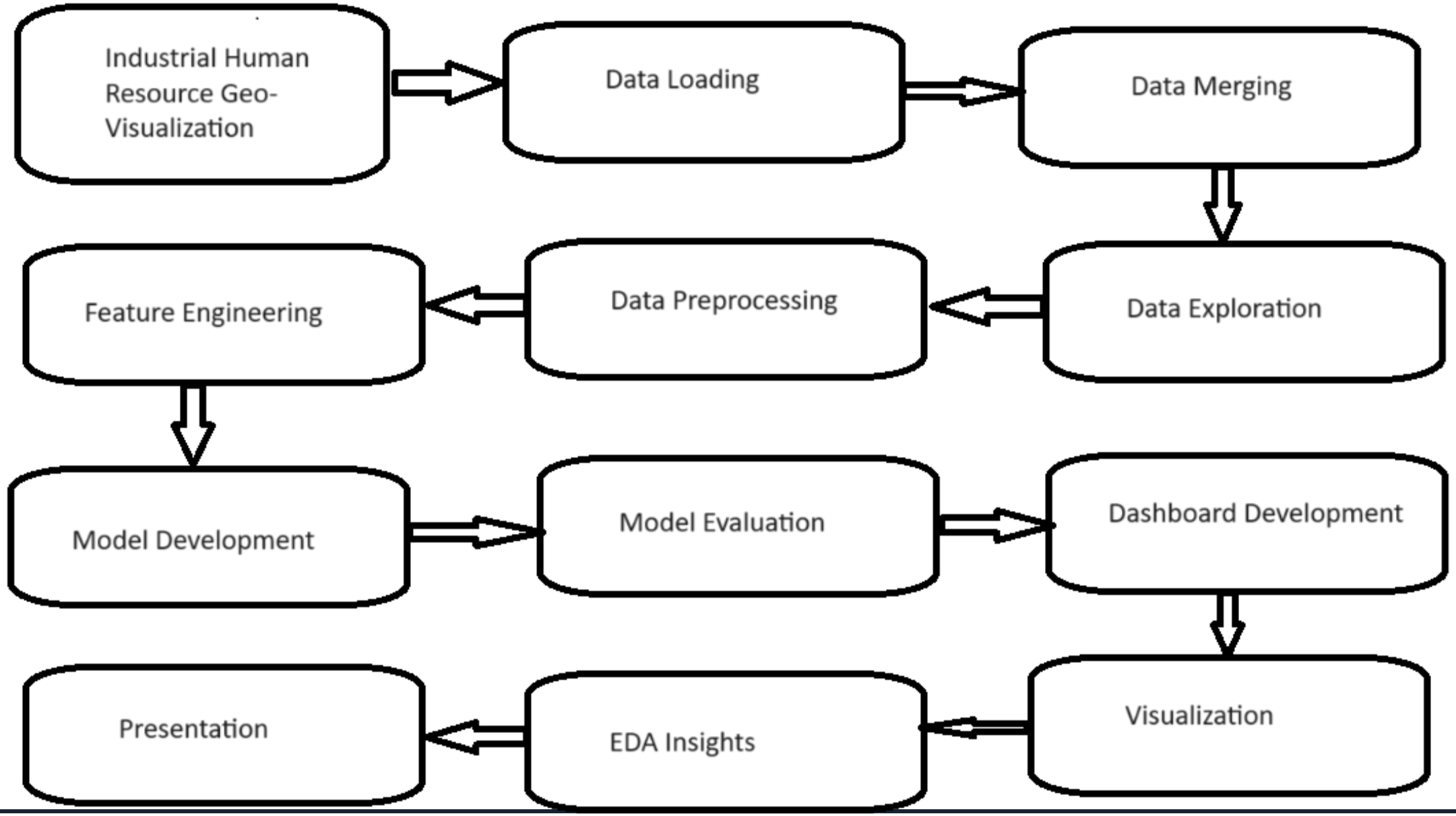
- The NLTK module is a massive tool kit, aimed at helping you with the entire Natural Language Processing (NLP) methodology.
  In order to install NLTK run the following commands in your terminal.

Stemming, lemmatization

- **Stemming** is a method in **text processing** that eliminates prefixes and suffixes from words, transforming them into their fundamental or root form, The main objective of stemming is to streamline and standardize words, enhancing the effectiveness of the **natural language processing** tasks.

- It looks beyond word reduction and considers a language's full vocabulary to apply a **morphological analysis** to words, aiming to remove inflectional endings only and to return the base or dictionary form of a word.

WORK FLOW


Industrial Human Resource Geo-Visualization
Data Loading
Data Merging
Data Exploration
Data Preprocessing
Feature Engineering
Model Development
Model Evaluation
Dashboard Development
Visualization
EDA Insights
Presentation

Conclusion:

Through the aforementioned approach, this project aims to address the issue of outdated and potentially inaccurate industrial classification data in India. By evaluating the existing dataset, identifying limitations, and leveraging machine learning and NLP techniques, the project seeks to provide updated and reliable information for policymakers and stakeholders. The development of a user-friendly dashboard will further facilitate data exploration and decision-making processes. Ultimately, this project contributes to enhancing the understanding of the Indian workforce's distribution across various sectors, thereby aiding in policy formulation and employment planning efforts.