Team ID: TM003070
Problem ID: PID380

Scheme Based Report Generation of Beneficiaries

Synopsis Abstract:

This plan is for a scheme that involves tracing the beneficiaries from the humongous database.

The database can separate into two forms for the various benefits available firstly in terms of gender (male & female). Each of the individuals will be then tracked based on the age and marital status. Different plans and benefits are supported for various tracks allotted based on the classification performed. For example, if the individuals are above 60 then they are eligible for schemes related to finance like BPL Income or National Old Age Pension Scheme. If the woman is termed as a widow, then they shall be teamed for the Widow Pension Scheme. If the criteria are not satisfied then the overall flow shall be declined or be accepted when the terms fit. This plan involves clearly categorizing and managing the schemes set for the citizens. The flow also allows the manager to sort and function different sects in a proper classification method.

Team Details: -

Name: - Karrtik Baheti
E-mail: - karrtikbaheti159@gmail.com
Mobile No.: 985656802
Institute: Ganpat University U.V. Patel College of Engineering
Semester:7
Branch: B. Tech (Computer Engineering - Artificial Intelligence)
Role: Team Leader

Literature Review/Existing Innovation-technology to address related to your problem:

We have thoroughly analyzed the problem and machine learning model has a solution to the problem.

We have already worked on different machine learning problems of classification and successfully build a machine learning model to counter the problem with top notch accuracy and precision, we are thorough with various algorithms used to solve the problems for the same.

Currently there is myScheme which asks for details and gives us the eligible schemes for a given person. Any student for reference can access this site and it will give out a list of schemes in categories i.e., for education, travel, social welfare, health etc and in that it gives out list of schemes with their respective links.

There is room for improvement here as our data is already there with government officials stored in database which can be used to get the list of available schemes for that citizen and we can notify them using a well-defined system, so that they can be aware of them and make full use of it. The platform helps the citizen to find the right Government schemes for them. It also guides on how to apply for different Government schemes. Thus, no need to visit multiple Government websites.

myScheme platform is Developed, Managed, and Operated by National e-Governance Division (NeGD), with the Support of Ministry of Electronics and Information Technology (MeitY), Department of Administrative Reforms and Public Grievance (DARPG) and in partnership with other Central and State Ministries/Departments.

Team ID: TM003070
Problem ID: PID380

What would be your approach to solve the problem:

Steps:

A. Creating a machine learning model
B. Creating a powerful backend structure which can handle API related tasks for analyzing the details and providing Authentication for admin and individual
C. Creating a frontend with reactjs for Interactive use.

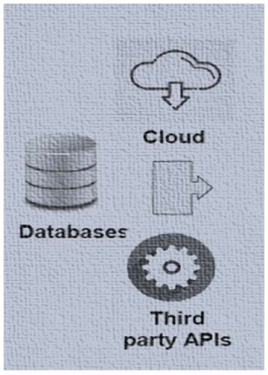
Steps for machine learning model:

1. Data Aggregation:

   Collecting data from various sources provided internally among government organizations which will be trustworthy.

   Database are constantly updating for every applicant around the country because of the Digital India initiative which will be in our best interest.

   There will be a big process involved in (re)Designing the data from various sources (or API's)



2. Data Analysis

   Data is a collection of information which will be useful for us throughout the process and will be base of the prediction. We will be analyzing the data thoroughly for the Age, Marital status, and dependents of the applicant with our collected data.

3. Data Pre-Processing

   Data generated from various sources will be containing a lot of noise and missing values around the generic data. There are some steps which will be helpful in overcoming this challenge like:

   a. Imputing numerical values with strategy like median to be filling values in place of missing data in numerical columns, and categorical values which will be ignored when the unknown values encounter in present in our training dataset to be able to predict the solution
   b. Many of the parameters extracted for variable selection depend on the data used for training will be reproducible.
   c. Calculating statistical values like the mean, median or mode can be helpful when there is a need of replacement for missing value.

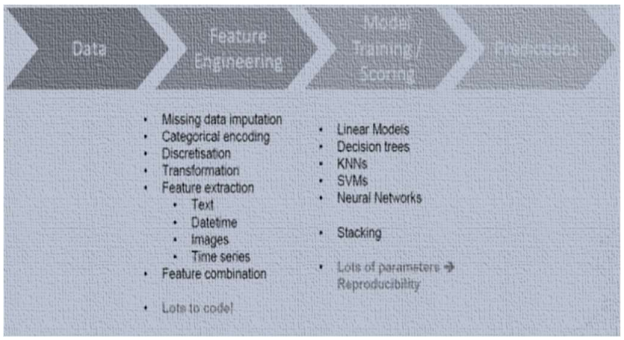
4. Variable Selection

   After we have efficiently completed the pre-processing step and handled the noise in the data here, we will select the different features from the given data.
   Features are the data items which are co-related with the target and we need to choose those features with maximum relation with the target.

   We will use the covariance matrix to check the co-relation between various features and the target.
   Covariance matrix will give the values of co-relation as output and the analysis of the values will revert us with useful features from the data to get the accurate output.

5. Machine Learning Model building

Team ID: TM003070
Problem ID: PID380

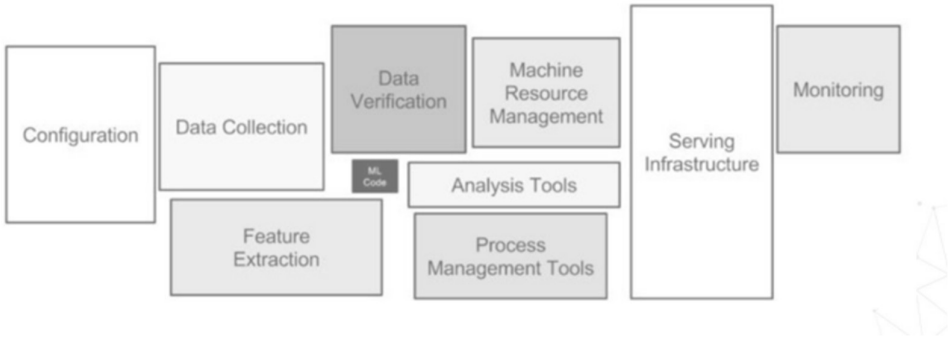


We had to make a choice between regression and classification, we opted for classification as targets of the problem can easily classified into different classes and can be predicted.

Another reason to go for classification is that the data elements would be static groups of point when plotted on a graph unlike regression were the data points are continuous when plotted on the graph.

6. Model deployment:

Integration of Machin learning model with your framework is a difficult task around which your deployment will be based as there is a difference between research and production environment such as configuration issues, Data dependencies, Data and Feature Preparation and scaling challenges which will be a major task to overcome in this whole situation for this type of huge organization.There will be always some updates in policies and development also requires constant monitoring for the major processes.



After working with ML model there is a need to integrate with our website based on ReactJS framework which will be fully able to handle individual request and batch request to process.

Tools and technologies to be used to solve the problem:

Tools: Django framework, Django REST framework, React JS, Python, Jupyter Notebook

Libraries: SkLearn, Pandas, Numpy, Tenserflow, Keras, jsonify

Technologies: Machine learning model, Python 3.7.8, React JS

Challenges/Risk in implementing your Final prototype (Risk):

Data format will be a challenging task for us as it all depends upon the raw data which will be brought to us by government organizations in different structure, format and it will be repetitive and sometimes generating a conflict in our data.

Possible outcome of your work:

It will be used for an individual or batch depending on who is using that website at that time.

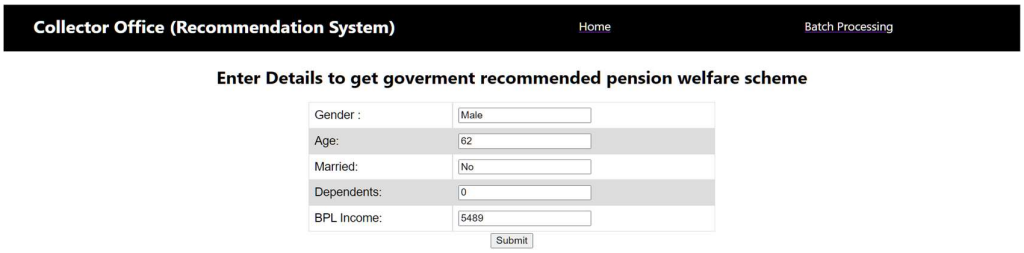
As for admin users we can provide them functionality like batch processing and for pension users we can provide them the necessary information on their fingertips

Work done till date:

Working ML Model applied with in django in a form where it can be used as an **API** and website is up and running on a local server just, we are waiting to be ready to be comply with the database provided from in whichever compatible format

Team ID: TM003070
Problem ID: PID380

Image/Screenshot of Solution (1):



Image/Screenshot of Solution (2):