Optimizing Vocational Training Budget Allocation to Minimize Unemployment in India

Krishna Jain (BSD-DH-2415)
Rahul Pandit(BSD-DH-2420)
Tanish Kothari(BSD-DH-2424)

Indian Statistical Institute, Delhi
Bachelor of Statistical Data Science

September 25, 2025

Problem

- Unemployment is a major socio-economic issue in India, affecting millions across both rural and urban regions.The problem is particularly acute among marginalized communities (SC/ST/OBC) and in less developed states with limited access to skill development and job opportunities.
- The primary challenge is to ensure equitable and efficient distribution of this budget to achieve maximum impact.
- This project aims to design an optimization model for budget allocation. Allocate funds across different states, sectors (rural/urban), and social groups (SC, ST, OBC, Others). Maximize employment uplift using real data from PLFS data and ensuring that allocation is data-driven and aligned with social equity mandates.

Data Extraction

The Periodic Labour Force Survey (PLFS)(REF: FROM MOSPI WEBSITE) provides detailed microdata on employment and household economic conditions across India.

- **Household-level data (CHHV1.txt):** Contains household characteristics including household size, social group (caste), and household consumption expenditure.

  It include data like caste, number of member, expenditure, vocational training, social group, state, household type, etc
- **Person-level data (CPERV1.txt):** Contains individual employment status and demographic details.

  It include data like gender, education level, employment status, monthly, income, etc
- **Data layout file: Data LayoutPLFS Calendar 2024.xlsx:** specifying variable names and formats.

Modeling Approach

The core objective of the model is to maximize potential employment gains by allocating training funds to groups with the higher need and unemployment burden. Each group is defined as a combination of state ($i \in [1, 37]$), sector ($j \in [1, 2]$), and social group ($k \in [1, 4]$).

- **Employment uplift proxy** tells us which groups would benefits most from extra budget to boost employment, by considering both there joblessness and their monthly expenditure:

$$emp\_uplift\_proxy_{i,j,k} = 0.7 \cdot unemp\_rate_{i,j,k} + 0.3 \cdot need_{i,j,k}$$

- The **need** is defined based on average household consumption expenditure (HCE_TOT) as:

$$need_{i,j,k} = \frac{1}{\text{mean HCE\_TOT}_{i,j,k} + 1}$$

  This ensures that poorer groups, who likely have lower consumption, are prioritized in the allocation.

- The overall priority score for each group is then computed as:

$$e_{i,j,k} = emp\_uplift\_proxy_{i,j,k} \cdot total\_population_{i,j,k}$$

- We normalize these scores to determine allocation proportions:

$$E_{i,j,k} = \frac{e_{i,j,k}}{\sum e_{i,j,k}}$$

Optimization Formulation

We define a linear programming (LP) problem:

$$Maximize : \sum_i \sum_j \sum_k E_{i,j,k} X i, j, k$$

- **Total Budget Constraint:** The total allocated funds cannot exceed the available budget(the Government of India has allocated a vocational training budget of **20,000 crores**(REF: THE HINDUS ARTICLE 2/2/2025) in the Union Budget 2024-25):

$$\sum_i \sum_j \sum_k Xi, j, k \leq B$$

- **Social Equity Constraints:**
  - At least 15% of the budget must be allocated to SC grousp.

$$\sum_i \sum_j Xi, j, 1 \geq 0.15B$$

  - At least 7.5% must go to ST groups.

$$\sum_i \sum_j Xi, j, 2 \geq 0.075B$$

- **State Equity Constraint:** Each state should receive a minimum of 0.1% of the total budget to ensure inclusion.

$$\sum_j \sum_k Xi, j, k \geq 0.001B, \forall i \in [1, 37]$$

- **Rural Equity Constraint:** Each state should receive a minimum of 10% of the total budget to ensure inclusion.

$$\sum_i \sum_k Xi, 1, k \geq 0.1B$$

These constraints help maintain a balance between efficiency and equity, ensuring that all regions and communities receive a fair share while still prioritizing those in greatest need.

Implementation

- Employment status was based on occupation codes. For each group, totals were calculated by population, employment, and spending. Then, an employment uplift measure was worked out for each group.
- Using the `lpSolve` package in R, we constructed and solved a linear programming problem with the above constraints. The resulting allocations were exported and analyzed by state, sector, and social group.

	ST	SEC	SG	total_population	total_employed	avg_expenditure	emp_rate	unemp_rate	need_exp_proxy	emp_uplift_proxy	unemp_priority	ab2
1	1	1	1	1357	573	16693.191	0.4222550	0.5777450	5.990108e-05	0.4044395	0.6	559164381.7
2	1	1	2	693	279	17181.866	0.4025974	0.5974026	5.819751e-05	0.4181993	0.6	295272235.5
3	1	1	3	1349	457	11977.718	0.3387695	0.6612305	8.348139e-05	0.4628864	0.7	742231328.8
4	1	1	9	5273	1787	15053.082	0.3388963	0.6611037	6.642717e-05	0.4627925	0.7	2900661250.0
5	1	2	1	101	27	23271.851	0.2673267	0.7326733	4.296852e-05	0.5128842	0.8	70369669.5
6	1	2	2	538	198	20304.279	0.3680297	0.6319703	4.924828e-05	0.4423940	0.7	282907427.9

Figure 1: Snipshot of Resulting allocation data

Findings and Analysis

The model produced an optimal budget distribution that reflected both economic deprivation and unemployment levels. Key findings include:

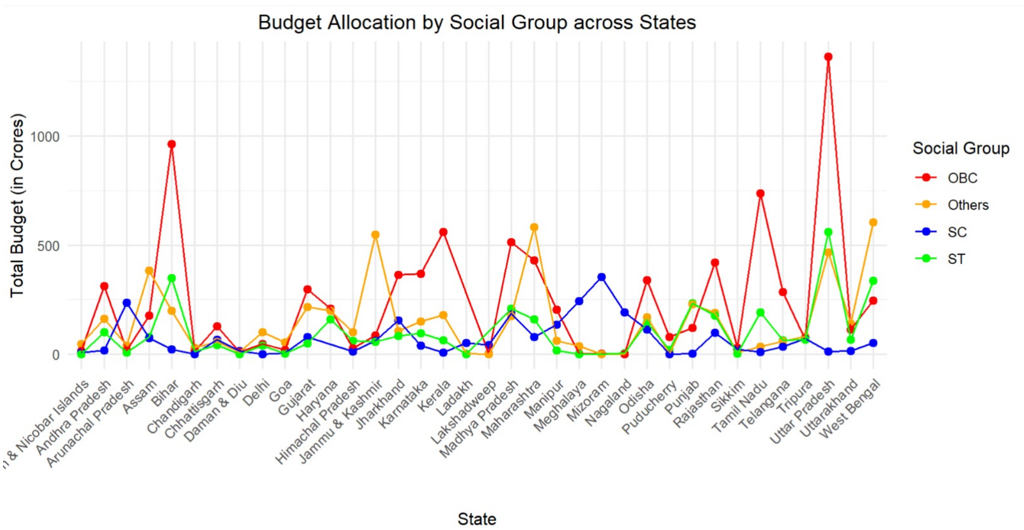

Figure 2: Optimal budget allocation by social group

- OBCs receive the highest allocations in most states, with major peaks in Bihar, Uttar Pradesh, and West Bengal (over 1000 crores).

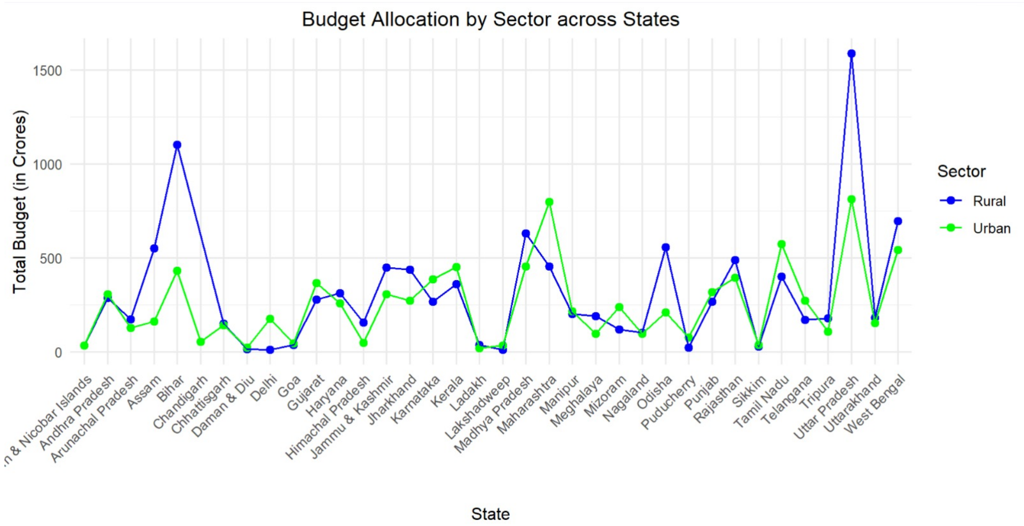

Figure 3: Optimal budget allocation to rural and urban

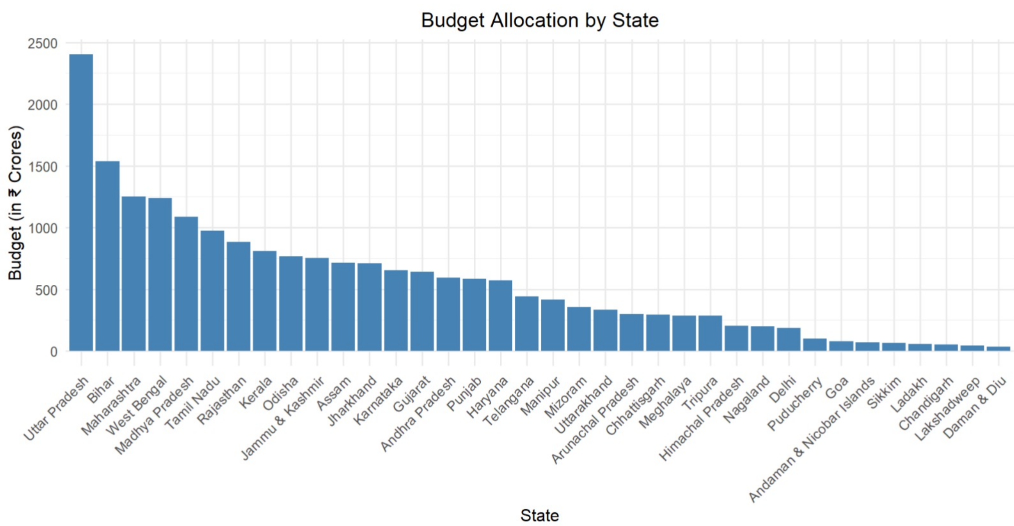

Figure 4: Optimal budget allocation to each state

- Significant variation across states: Larger states get much higher allocations; smaller states/UTs get less.
- SC and ST allocations are lower but consistent, with relatively higher shares in states where these groups are populous (e.g., Chhattisgarh, Jharkhand, Odisha).
- The “Others” group gets moderate allocations, occasionally spiking in states like Maharashtra and Rajasthan.
- Some sharp spikes for OBCs suggest areas for further investigation (demographics, unemployment, or model parameters).

- Budget Allocation by Sector across States Rural sectors receive higher allocations in most states, especially in Bihar, UP, West Bengal, and Assam.
- Urban sectors receive significant allocations in states like Maharashtra, Gujarat, and Delhi.
- There is wide variation in total allocations by sector and state, reflecting population and need.
- Both rural and urban sectors receive funding in every state, reflecting balanced geographic equity.

Limitations

Despite its strengths, the model has several limitations:

- The use of proxy variables (e.g., HCE_TOT for economic need) may not fully capture the multidimensional aspects of deprivation.
- The model assumes uniform effectiveness of vocational training across regions, which may not hold true in practice.
- The model is static, representing a one-time allocation instead of a dynamic, multi-period policy that adapts to evolving economic realities.
- It does not account for the capacity of states or regions to absorb and utilize funds effectively.

Conclusion and Future Work

This study demonstrates how optimization techniques can be used to guide equitable and effective allocation of government budgets. By combining unemployment data with economic need indicators, the model ensures that funds are directed toward areas with the highest potential impact.

In future iterations, the model can be enhanced by:

- Incorporating additional demographic variables like age, gender, and education levels.
- Including skill gap analysis and regional training infrastructure data.
- Making the model dynamic to support multi-year budget planning.
- Integrating training effectiveness data to estimate employment outcomes more accurately.

Overall, this framework provides a transparent and data-driven approach to reducing unemployment through targeted vocational training initiatives.