

# Union budget analysis of India using R

<sup>1</sup>Swathi Harish, <sup>2</sup>Swati N Murthy, <sup>3</sup>Ashwini B P

<sup>1,2</sup>Student, <sup>3</sup>Assistant professor  
Computer Science of Engineering,  
Siddaganga Institute of Technology, Tumakuru, India

**Abstract**— The Union Budget is the annual financial report of India. It contains the government of India's revenue and expenditure for the end of a particular fiscal year, which runs from April 1 to March 31. Some of the problems faced during the planning of budget are: Extra expenses spent on supplementary budget, uneven distribution of money to departments. Since there is a limit to the resources the need for proper budgeting arises to allocate scarce resources to various government activities. And also, manual handling of the budget leads to human error. The proposed system will build software that will predict if the current proposed budget will be successful or not. It will take the previous 10 years' budget as the input and do certain analysis upon it and predict the result.

**Index Terms**—Implementation budget, Expenditure budget, Economic state, Classification, Prediction.

## I. INTRODUCTION

There are many factors that influence the progress of a nation. Steady government, food availability, small scale and large scale industries, export and import are some of such factors. But all these factors stand on a strong foundation called economy of the nation. A stable economic condition can be established only when proper planning is carried out. This planning is done every year by the government in the name of the budget declaration.

The Union budget of India, also referred to as the Annual financial statement in the Article 112 of the Constitution of India, is the annual budget of the Republic of India. It is presented each year on the last working day of February by the Finance Minister of India in Parliament. This budget is efficiently planned by a group of highly qualified officers from many days of continuous effort. The Finance ministry also provides an account of implementation details about the execution of plans mentioned in the budget, this is declared as an Implementation budget every year. Planning and declaration of the supplementary budget requires an extra set of efforts and time. It would have been much simpler if the allocation of money was right in the first budget itself. The main objective is to introduce advanced technology in the field of governance and help in the reduction of manual efforts to a maximum extent.

### B. Proposed system

Figure 1 shows the architecture diagram of the proposed system. The application can be accessed by everyone over the internet using a URL. Many users can access the application at the same time. The application is hosted on a cloud platform named shinyapps.io. The application makes use of the shiny package in R to construct a web application. The shiny package of R consists of the files that shows the user interface and the file that runs the server side functionalities in response to the user actions performed.

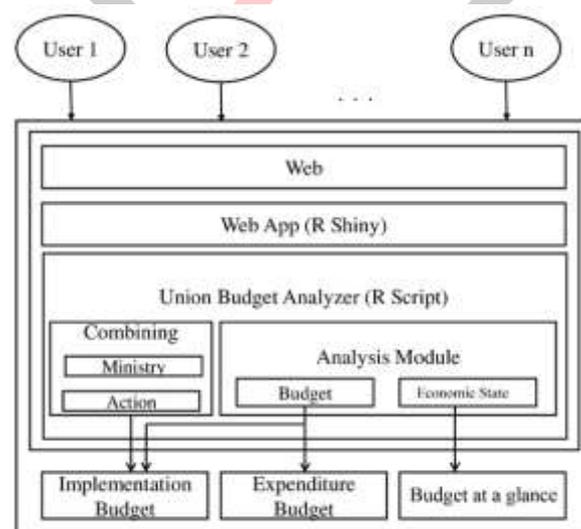


Figure 1

Rshiny makes use of the R files to run the application. Different R files are written to perform budget analysis and the receipts from plans analysis. These R files are run(executed) upon click of a button in the shiny web interface. The R files that are run in the background include the budget data analysis files and the budget at glance files. Lots of pre-processing is done over the data. Data

pre-processing is a data mining technique that involves transforming raw data into an understandable format. The data is often inconsistent, consists of spurious tuples, values, like extra spaces, extra lines, non-readable characters, among others. Data preprocessing prepares raw data for further processing. The data preparation and preprocessing thus takes up a lot of time. The outcome of the data pre-processing step is called the final training set.

The training set obtained after pre-processing is taken as an input to the analysis stage. Here, mathematical transformations should be done and certain predictions are made. For prediction and analysis, in case of union budget, the past 10 years' budget is taken as the reference data set.

Graphs, Charts like pie chart, bar graph and line graphs are produced using the output obtained.

### C. Organization of the paper

The rest of the paper is organized as follows: section 2 gives the related work, section 3 presents the proposed system, section 4 concludes the paper, section 5 presents our heartfelt acknowledgments.

## II. RELATED WORKS

*India's Union Budget: Changing Scope and the Evolving Content* [1]: assesses the contents of union budget and the finance minister's speech. It also reflects the evaluation criterion for budget contents, processes involved and suggests some changes in fiscal policy and underlying budget processes.

*R: A Language for Data Analysis and Graphics* [2]: Ross Ihaka and Robert Gentleman, the authors of the R language, have shared their experience in designing and implementing a statistical computing language are discussed. In developing this new language, useful features from two existing computer languages, S and Scheme were combined. The details about using the R language are mentioned in the paper.

*Getting started with Rstudio* [3]: discusses about the user friendly integrated development environment RStudio. The main complaint regarding R before the development of RStudio was that it was not friendly and multiple interfaces exist and differ across platforms.

*Implementation budget* [4]: gives the details of all the plans declared in the budget, their descriptions, details of the department(s) that carry out the plan and also the status of implementation.

*Expenditure budget* [5]: is a document that contains all the departments of the budget, their expenditure on plans and non-plans of the current year, revised expenditure of previous year and expenditure of the year before that.

*Budget at a glance* [6]: is a file that consists of various fields out of which a few are extracted. The important entries are against Receipts, Expenditures, Fiscal deficit, GDP, Growth percentage.

## III. PROPOSED SYSTEM: UNION BUDGET ANALYZER

Figure 2 shows the overall flow of the proposed system. Required parameters are as described in the algorithm:

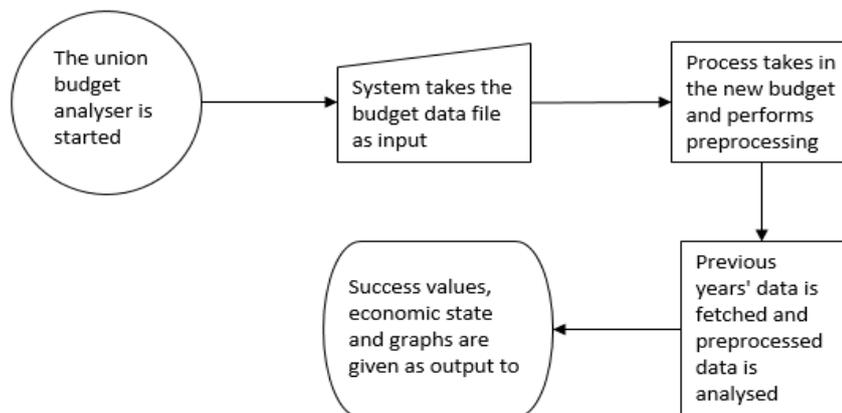


Figure 2

### Algorithm BudgetAnalyzer

Input:

choice = input\_from\_user

if choice is success\_predict then

    num = number\_of\_years = input\_from\_user

    while num do

        name = str(num)

        table1 = preprocess\_implementation\_data(name)

        table2 = preprocess\_expenditure\_data(name)

        table3 = combine\_data\_calculate (table1, table2)

        table4\_row = final\_data(table3)

        table4.append(table4\_row)

```

end while
res_analysis = classification_algorithm(table4)
weighted_list = assign_weight(res_analysis)
new_budget = input_from_user
table1 = preprocess_implementation_data(new_budget)
table2 = preprocess_expenditure_data(new_budget)
table3 = combine_data_calculate (table1, table2)
success_value = compare_predict_success (table3, weighted_list)
output_to_user(success_value)
else if choice is economic_state_analysis then
num = number_of_years
while num do
name = str(num)
table1 = preprocess_budgetAtAGlance_data(name)
table2 = preprocess_expenditure_data(name)
data = extract_essential_features(name)
table2.append(data)
end while
economic_state = get_state(table2)
output_to_user(economic_state)
end if

procedure preprocess_implementation_data(name)
vector1 = get_ministry(name)
vector2 = get_status(name)
weights = assign_weights(vector2,0,0.33,0.66,1)
table1 = combine (vector1, vector2, weights)
return table1
end procedure

procedure preprocess_expenditure_data(name)
vector1 = get_ministry(name)
vector2 = get_non_plan(name)
vector3 = get_plan(name)
table2 = combine (vector1, vector2, vector3)
return table2
end procedure

procedure combine_data_calculate (table1, table2)
name_of_dept = get_col1(table2)
contribution = number_of_plans(table1)/sum_of(number_plans_budget(table1)
num_of_plans = number_of_plans(table1)
weightage_of_plans = sum_of(get_row3(table1))
final_weight = weightage_of_plans / num_of_plans
table3 = combine (name_of_dept, contribution, num_of_plans, weightage_of_plans, final_weight)
return table3
end procedure

procedure final_data(table3)
vector1 = cnt++
c=0
for dept_name in table3.get_col1
vector(c++) = dept_name
department = contribution(table3) * final_weight(table3)
success_val = get_col4(table3) / number_of_plans(table3)
table4.combine(vector (x for x in c))
table4.append(department)
return table4
end procedure

```

*Vector1*: Row: Name of the Departments.

The Implementation budget is taken as an input in a text format. Originally it will be in PDF format and must be converted to the text format. It can be done using any online tool, we have used [convertio.co](http://convertio.co) [4]. Ministries are read and stored in a file using

regular expressions. The pattern that is used to extract data is that the ministries/departments begin with “Department of” or “Ministry of” or “D/o” or “M/o”.

*Vector 2: Row: Status of Implementation*

The Implementation budget contains the status of Implementation of a particular plan along with the nodal ministry or department associated with it. It consists of entries like, “Action Completed”, “Action partially completed”, “Work in progress”, “Ongoing process”, etc. Even this data is extracted with the help of regular expressions and stored in a file.

*Tables formed*

Table1 shows the data extracted from implementation budget. The department and status details are combined onto a single data frame for future analysis. data.frame function combines two vector or data frames only if they are of equal length. If the length of ministry and status vectors differ by a large margin, the processes carried out to obtain Vector1 and Vector2 are repeated again and the process is continued. A weight is assigned to each row based on the status of implementation achieved.

Table 1: The Implementation budget data extraction

Name of the Ministry	Status of implementation	Weight
----------------------	--------------------------	--------

Table2 shows the data extracted from expenditure budget. The expenditure budget is converted into text format and is taken as the input. The names of the departments, Planned expenditure, Non-plan expenditure for the current year are extracted into a data frame by using regular expressions. The revenue and capital rows are not considered.

Table 2: The Expenditure budget data extraction

Name of Department	Planned expenditure	Non-plan expenditure
--------------------	---------------------	----------------------

Table3 combines the data of both Table1 and Table2. It is as shown below. The joining criterion is the name of the department.

Table 3: Data combiner

Name of the Department	Contribution	Number of plans	Weightage of the plans	Final weight
------------------------	--------------	-----------------	------------------------	--------------

Once the implementation data and expenditure data has been pre-processed and formed into data frames, the total number of plans for each department is collected and made as a separate column under expenditure data as per Equation 1.

$$Number\ of\ plans(x) = Total\ plans\ in\ the\ department(x) \tag{1}$$

The contribution of each department is determined as the fraction of plans it has amongst all the plans. It is calculated as shown in equation 2.

$$Contribution(x) = \frac{Total\ plans\ in\ dept(x)}{Total\ plans\ in\ the\ budget} \tag{2}$$

The final weight for each department is also calculated by using the weight attribute of implementation data as in Equation 3.

$$Weightage\ of\ the\ plans(x) = \sum(Weight\ in\ implementation(x)) \tag{3}$$

Weights of all the plans coming under a particular department are added to form weightage of plans data under the expenditure budget. This is divided by the total number of plans to obtain Final Weight as in Equation 4.

$$Final\ weight(x) = \frac{Weightage\ of\ plans(x)}{Number\ of\ plans(x)} \tag{4}$$

The contribution of a department is calculated and stored as a separate column under the expenditure data by taking the percentage of plans of each department in the budget.

Table 4 gives the value of the contribution of each department to the success of the budget for that year. The last column gives the overall success of the budget for that year.

Table 4: The final table

Years	Department 1	Department 2	Department 3	. . .	Department n	Success Values
-------	--------------	--------------	--------------	-------	--------------	----------------

After this step of calculation is performed, the actual data for each department is obtained by multiplying the weight under expenditure budget by the contribution attribute. This is saved as the value for each department as in Eq.5.

$$\text{Department}(y) = \text{Contribution}(x) * \text{Final Weight}(x) \quad (5)$$

The last column under the final table gives the success of the budget. This is calculated for all ten years. It is calculated as shown in Eq. 6. Now the table is ready for analysis algorithm to be applied on it.

$$\text{Success of budget}(x) = \frac{\sum \text{Weight in implementation}(x)}{\sum \text{Number of plans}(x)} \quad (6)$$

#### IV. CONCLUSION

To reiterate, the union budget analyzer makes sure that the budget released considers the current requirements of a department and helps in proper distribution of wealth. In future the analysis can be done considering more factors and obtain more accurate results. It can also be extended with some modifications to college environments where budgets are allocated to each department, also to hospitals, banks and many such areas.

#### V. ACKNOWLEDGMENT

Firstly, we<sup>1,2</sup> thank our guide Mrs. Ashwini B P<sup>3</sup> for her support and valuable guidance throughout the duration of the project. We thank Mr. V Ananth Krishna, Retired Government official in the Finance Department of India and <sup>1</sup>Swathi Harish's grandfather, who introduced us to the budget analysis with his valuable ideas. We also thank the Finance ministry of India who made the budget files easily accessible. We are thankful to the developers of R and Rstudio without whom the dissertation work would have been very difficult. We thank everyone who contributed in the success of this work.

#### REFERENCES

- [1] R. Malhotra, *India's Union Budget: Changing Scope and the Evolving Content*, 2010
- [2] R. Ihaka and R. Gentleman, "R: a language for data analysis and graphics", *Journal of computational and graphical statistics*, vol. 5, no. 3, pp. 299-314, 1996.
- [3] J. Verzani, *Getting started with RStudio*. O'Reilly Media, Inc., 2011.
- [4] Finance Ministry, *Implementation Budget of India*. Government of India, 2015.
- [5] Finance Ministry, *Expenditure Budget of India*. Government of India, 2015.
- [6] Finance Ministry, *Budget at a glance*, Government of India, 2015.