Project Report

On

**Data4Democracy**

By

Shreyas Patil

Souryaansh Agarwal

Gopesh Khandelwal

Pranav Taneja

Under the supervision of

**Prof. Dipanjan Chakraborty**

SUBMITTED IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS OF

**CS F441: Selected Topics in Computer Science**



**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE PILANI (RAJASTHAN)**

**HYDERABAD CAMPUS**

**(FEBRUARY 2020**)

**ACKNOWLEDGMENTS**

Foremost, we would like to express our sincere gratitude to Professor Dipanjan Chakraborty for providing us with the opportunity to work on this project (Data4Democracy), and for his patient guidance and valuable support.

****

**Birla Institute of Technology and Science-Pilani,**

**Hyderabad Campus**

**Project Details**

**Duration:** 1st February 2020 to 10th July 2020

**Date of Submission:** 11th July 2020

**Title of the Project:** Data4Democracy: Comprehensive Rating System for MPs in Indian Parliament.

**ID No. Name(s) of student(s)**

2017A7PS1568H Shreyas Shashank Patil

2017A7PS1578H Souryaansh Agarwal

2016A7PS0101H Gopesh Khandelwal

2016A7PS0096H Pranav Taneja

**ABSTRACT**

This report describes our project on Data4Democracy. This project deals with developing a platform which can be used by voters to get information on Indian Members of Parliament (MPs). A comprehensive rating system has been developed so that MPs can rated on specific parameters, which would make it easier for a voter to get a perspective of the MP. This system has been implemented through a JavaScript-based website. This report explains how data was obtained, how the rating system was developed and how it was applied.

**Index**

|  |  |  |
| --- | --- | --- |
|  |  |  |
| Sr. No. | Title | Pg. No |
| 1 | Introduction | 6 |
| 2 | Sources and Literature Review | 7 |
| 3 | Parameters | 8 |
| 4 | Justification and Rubric | 9 |
| 5 | Web Scraping | 11 |
| 6 | Implementation | 12 |
| 7 | Limitations and Recommendation | 13 |
| 8 | Conclusion | 14 |
| 9 | Appendix | 15 |

**Introduction**

Lack of organised information about candidates leads to voters making uninformed choices during elections.

There does not exist any comprehensive framework or platform which people can use to get compendious information on the Members of Parliament.

There is thus a need for authentic information and analysis of every elected representative and candidate. The information should also be easily accessible, reliable and performance based.

There was a need to develop a rational, reasonable, fair, transparent performance-parameter assessment scale and apply it uniformly to all relevant data.

As a viable solution to this, we created a Performance Management System that uses multiple parameters to rate incumbent MPs, based on their efficiency and effectiveness, and presents the findings and analysis in a structured manner.

**Solution Sketch**

The basic concept was to create a platform which can be accessed by voters to get information on sitting MPs.

We planned to create a Performance Management System that uses multiple parameters to rate incumbent MPs, based on their efficiency and effectiveness.

Solution Outline

* We developed a platform (Web-based).
* Obtained relevant information using web scraping techniques.
* Our platform rates MPs based on key performance indicators.
* Ratings is based on a well-designed system of ‘performance management’.

**Sources and Literature Review**

There wasn’t any single source which contained all relevant data. Thus, we had to refer to multiple sources to compile all requisite data.

All factual information was obtained from the following sites:

* https://www.prsindia.org/mptrack/
* https://adrindia.org/
* https://www.mplads.gov.in/mplads/AuthenticatedPages/Reports/Citizen/rptDetailsSummary.aspx
* https://www.myneta.info/

**Review** of the past (and ongoing) work done in this domain:

The above mentioned sites, namely prsindia.org and adrindia.org, are open data depository platforms having an extensive collection of data on which they make analysis. Similarly, myneta.org provides voters with election related information, mainly dealing with non-quantitative attributes of MPs like education details, criminal records, finances and income-expenditure statements. Although, these platforms do not provide a direct rating of the MP based on all factors.

Outside India, there are a few sites which do a similar analysis, <https://www.change.org/l/uk/the-people-power-index-how-did-your-mp-score> is one such site which analyses data from the UK Parliament and provides results and suggestions for improving efficiency, although it considers a limited number of parameters for this purpose, and thus falls short of providing a fair and thorough rating system.   
Similarly, <https://www.theguardian.com/environment/ng-interactive/2019/oct/11/guardian-climate-score-how-did-your-mp-do> analyses MPs work in a very niche domain, i.e. climate change, and provides ratings and results. But this is a very specific rating methodology, and would not be very useful for people to have a complete idea about their MP.

**Parameters used for Analysis and Rating**

MPs have been rated based on the following parameters:

1. **Quantitative Parameters**:

● Attendance

● Participation in Debates

● Number of questions asked

● Bills proposed

● MPLAD spending pattern

● Membership in Parliamentary Committees

● User Ratings

(These parameters have been used to rate MPs on a uniform linear scale.)

2. **Qualitative Parameters**:

● Educational Qualification

● Assets and Liabilities

● Criminal Record

For qualitative analysis, Likert Scale has been used.

**Justification and Rubric**

Justification for the parameters chosen:

* Attendance - Your MP's job is to represent you, he is your voice in parliament. This looks at how many working days an MP attended the parliament.
* Debates - It shows their active participation on various issues in parliament. Higher number of debate participation shows that the MP is active in multiple areas and makes his voice heard, thus representing the interests of his constituency better.
* Questions - It is an MP’s way to get to know the impact of national policies on his constituency. Higher scores indicate that the MP likes to raise issues and questions to ensure that the nation or his constituency does not get adversely affected by a new law or topic currently in discussion.
* Bills proposed - It shows initiative taken by your MP in drafting innovative policies.
* MPLAD spending - It helps you gauge if your MP has spent the funds wisely on pressing issues in your constituency.
* User Ratings - Shows the current user opinions and sentiments.
* Membership in Parliamentary Committees - This is how an MP gets involved with influencing policy. Tells you what issues your MP cares about.
* Number of Terms – Higher number of terms implies that the majority of people in the particular constituency trusts the MP.

Non-quantitative parameters:

Background Info - Education, Court cases, assets and liabilities.

Parameter-wise rubric for rating:

1. **Performance Ratings**:

* Attendance: Percentage value is taken, and then scaled down to 30%.
* Number of Debates: Numeric value scaled up to 25% (based on mean).
* Number of Questions asked: Numeric value scaled up to 30% (based on mean).
* Number of Private Member Bills introduced: Numeric value taken, converted to percentage, then scaled down to 15%.

**NOTE**: Ratings for debates, questions and private bills are relative, because there is no absolute standard or pre-existing marking system.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Debates | Score | Scaled | Questions | Score | Scaled |
| 0 | 0 | 0 | 0 | 0 | 0 |
| 1-2 | 1 | 2.5 | 1-4 | 1 | 3 |
| 3-5 | 2 | 5 | 5-10 | 2 | 6 |
| 5-8 | 3 | 7.5 | 11-15 | 3 | 9 |
| 8-12 | 4 | 10 | 16-30 | 4 | 12 |
| 13-20 | 5 | 12.5 | 31-45 | 5 | 15 |
| 21-30 | 6 | 15 | 46-60 | 6 | 18 |
| 31-50 | 7 | 17.5 | 61-75 | 7 | 21 |
| 51-75 | 8 | 20 | 76-100 | 8 | 24 |
| 76-100 | 9 | 22.5 | 101-150 | 9 | 27 |
| 100+ | 10 | 25 | 150+ | 10 | 30 |

National Debates average: 13.6  
National Questions average: 38.0  
National Attendance average: 85%

1. **Approval Ratings**:

* User Ratings: Average numeric value (obtained on a 10 point scale) scaled up to 50%.
* Number of terms: Numeric value scaled up to 50%.

**Likert Scale rubric for Qualitative parameters**:

A 3-point Likert scale has been chosen to mark various parameters:

* Education: High, Medium, Low.
* Number of Criminal Cases: High, Medium, Low. (Inverted Scale)
* Assets: High, Medium, Low. (Not directly quantifiable)

Education: High (University level (Graduate, Post graduate, Doctorate)), Medium (Matriculate, under matric, Diploma), Low (Inter/higher secondary).  
Criminal Cases: High (>4), Medium (1-4), Low (0).  
Assets: High (> 10000000), Medium (3000000-10000000), Low (<3000000).

**Web Scraping**

Since the required data was large, and complete readymade data was not available in one place, we used web scraping to obtain the required data, and compiled it into a CSV file.

* Web scraping was done using the Python library BeautifulSoup (for parsing structured data).
* Image downloading was automated using Pillow (python library) and Selenium.

**Implementation**

The entire rating system described above has been implemented as a website. The reason for this is that a website can be accessed by anyone with an internet connection, irrespective of the device and operating system they have (same thing cannot be said for other options, like android apps).

**Source code of the website can be obtained on this link:** [**https://github.com/gopeshkh1/sel\_topics\_assignment**](https://github.com/gopeshkh1/sel_topics_assignment)

**Technical Environment**Technical requirements to run this portal on the system:   
(Requirements to directly run this directly on the system; once the site goes online, only an internet connection and a browser is required to access it).

Node.js needs to be installed on the PC, flowing which the npm package needs to be installed. Once installed, we can directly run the client and the server.

1. Install the packages

npm install

1. To run the client for development

npm run client-start

1. To build the client side for distribution

npm run client-build

1. To run the node server for development

npm run server-start

**Website Structure**:

The site offers the following features:

* Dashboard
* MP search by name and by constituency
* Filters for State and Constituency
* Ranked list of MPs based on a particular parameter
* MP information
* Pagination
* Allows user to rate MPs

**Limitations and Recommendations**

General issues:

* The rating criteria that we use has been formulated after taking information and ideas from multiple sources, and we have tried to cover as many aspects as possible. Still, there are several subjective factors that are impossible to analyse completely (for example, the MP only has a recommendatory role in MPLAD spending, its implementation is controlled by the district authority).
* The ratings (both quantitative and qualitative) may or may not give a complete overview of the MP, depending on the information available (for example, for several MPs, education details are not available).
* There are a lot of factors which cannot be quantified, for example, how an MP listens to the public, how many of his pre-election promises does he fulfil etc. (We have added User Ratings for this purpose, wherein users can rate the MP based on these factors, but there is no way to guarantee the authenticity of this aspect).
* Implementation of User Rating would not be much useful unless we can control it, i.e. only people belonging to a certain constituency should be able to rate the MP of that area.

**Recommendations** (Upgrades that could be made in the future with the right amount of resources and skills):

* More comprehensive analysis: This dataset can be expanded to include analysis of previous MPs of each constituency, encompassing at least 3-4 terms. This would give a bigger dataset and would help people to compare MPs easily. This would also help us observe patterns and make better decisions.
* Prediction: Machine learning can be used on larger datasets. This would have dual benefits, it could predict ranks on its own, and it could also introduce suitable new rating parameters on its own.
* Sentiment Analysis: Quantification of an MP based on data leaves out certain factors like public opinion and likeability. Crowdsourcing this kind of data is likely to provide a better picture of the MP. One way to do this is to do a Sentiment Analysis from various platforms like Facebook and twitter.
* Solution to User Rating issue: We can add user authentication, so that one person can only vote once from his verified profile. Profile verification can be done by asking users to provide certain proofs, for example proof of residence.

**Conclusion**

This report describes our efforts to develop a platform to help voters make the right choice during election. We have done this by building a website which gathers data and analyses it. This is then used to rate MPs based on several quantitative and qualitative parameters.

Once deployed, this would help users gain a quick insight into the performance and life of the incumbent MP. The website is fully functional, although it is still in its Beta version. As suggested in the Recommendations section, a few upgrades would make this site much more useful.

**Appendix**

Source code: <https://github.com/gopeshkh1/sel_topics_assignment>

Screenshots of the working site: