

**Visvesvaraya Technological University,**

**Jnana Sangama, Belgaum - 590014**

A SRS project report on

Prediction of Electricity Generation for Better Decision Making

Submitted in partial evaluation of the 8th semester project progress review-1

Submitted by:

1PI13CS103 Parag Karguppikar

1PI13CS104 Parikshit Maheshwari

1PI13CS121 Rathan Gangadhar naik

Under the guidance of

**Prof. V R Badri Prasad**

**Jan – May 2017**

|  |  |
| --- | --- |
|  | **Department of Computer Science & Engineering**  **PES Institute of Technology,**  **(An Autonomous Institute under VTU, Belgaum and UGC, New Delhi)**  **100FT RING ROAD,BSK 3RD STAGE, BENGALURU - 560085** |

Abstract

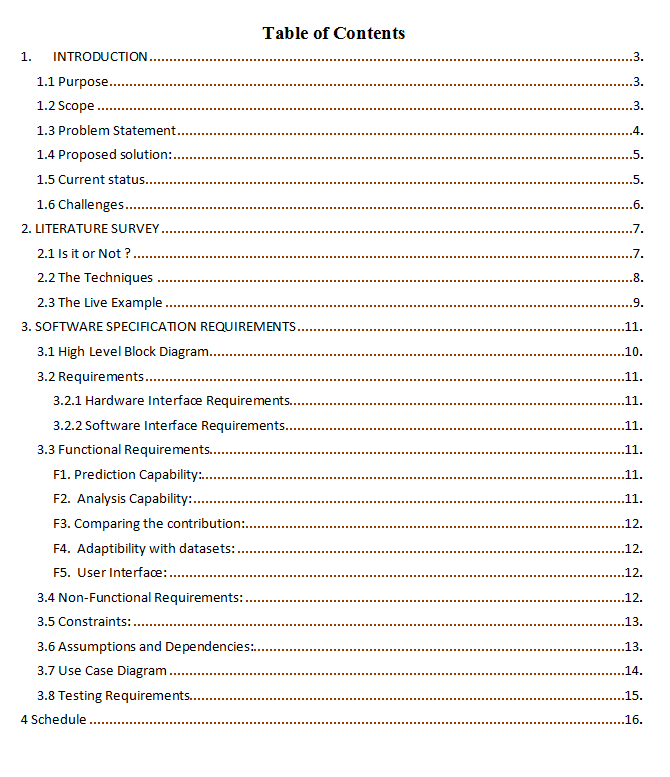
From the perspective of this planet’s clock, the time spent from a human realizing his humanity to the present, only account for one second. And within that one tiny second of arrival, the mankind has become dependent on infinite resources. Electricity happens to be one of the front-runners of one of these uncountable resources.

One of the biggest concern of our country is to supply electricity to 120 crore people. As India doesn’t have proper resources to generate such huge amount of electricity. We depend more on the imported coal, gas and diesel. And the government is trying its best to increase the generation of electricity from renewable resources like solar, wind etc.

Now the biggest challenge is, even though we generate so much power, the problem comes in its usage. How to use the generated power efficiently? The (Transmission and Distribution) T&D losses are around 22% which are not affordable. And due to this, it has become a very difficult task to provide electricity to all the states every day, especially in rural areas where infrastructure is very poor and T&D losses are more.

Our project here will try to understand how electricity is generated from various resources ranging from thermal to nuclear and in between involving hydro and various renewable resources. Here our main focus is on how India fares in electricity generation.

A nation of 28 states, the country is a witness of every major electric power plant working. Each of these major players contributes energy (GWh). We will try to assess and form a pattern out of these findings, which will help us in understanding the dynamics of the present situation and prepare for the future.



**Chapter - 1**

**Introduction**

* 1. **Purpose**

India is one of the country developing at the pace of a rocket reaching for the skies. Every day new industries get opened and thus increase in fulfilling the energy requirement of those industries. But here is the catch – the generation part of the electricity is more of a constant as compared to every increase in the rate of consumption.

The power sector is mainly divided into 3 parts. That is the generation, transmission and distribution. The generation is power is spread across 100’s of plants and then transmitted to either state grids or center grid. The pattern in which this works is very complex, as the generation depends on a lot of parameters which are not under human control like rain, temperature etc.

So there was a need for a system, which should guide our government and any other researchers in the field to handle the complex dynamics of electricity generation and if possible, even consumption. The system should predict the power generation of these individual plants and helps the relevant people to make decisions on distribution and transmission with minimal losses. Also, it should help in deciding the amount of fuel to be imported in order to fulfill the needs of industries and domestic usage.

The project will give different graph plots as output, which will show the generation of electricity by different states and for even different categories. It will also compare the generation of power over the years. This will give a pattern to predict the generation over the next few years accompanying different parameters.

Although we considered the major parameters for prediction, there are some more minor parameters which affect the generation like lockouts, maintenance issue etc. These parameters are not included in our project due to unavailability of data.

* 1. **Scope**

As from the data obtained from Open Government Data Platform India, every state produces electric energy through more than one power plant stations. These power plants are mainly categorized below along with their percentage contributions:

* Thermal Power Stations
  + Coal & Lignite (61 %)
  + Gas (8.2 %)
  + Diesel (0.3%)
* Hydro Power Stations (13.9 %)
* Renewable
* Small Hydro (1.4%)
* Wind Power (9.1%)
* Biomass (1.6 %)
* Solar Power (2.7%)
* Nuclear (1.9%)

These power stations have the pre-installed capacity for power generation (MW). For each month, an estimate GWh energy production is decided. At the end of that month, how the energy generated is compared with the pre-determined estimated value.

For example, take one state, say Uttar Pradesh. This state has 26 power plant stations as of December 31, 2016. Out of 26, 1 is nuclear, 4 are hydro and the rest are thermal power plant stations. Each of these power plant stations has their power generation installed capacity mentioned. Take one particular month from the last decade and it will say how much they are estimating out this station.

Now collecting all this data and analyzing through computing rather than manual study will give us a clearer picture on how each state contributes for the net electricity generation and combining this study with other parameters say rainfall for hydropower stations may determine the generation for upcoming months.

* 1. **Problem Statement**

Accountability in power generation is the biggest growing concern in India’s power sector. Managing the fuel imports, their usage, distribution to various sectors is always unplanned and done on the basis of the amount of electricity generated that year or month.

Our project will try to understand the factors affecting the electricity generation of major power stations (thermal, hydro and nuclear) and predict the amount of electricity that is going to be generated next year by each sector. Also give an analysis of the generation in various states and their dependency on the factors affecting generation.

* 1. **Generic Proposed Solution**

The proposed solution is to extract the meaningful parameters from the datasets which have the estimated, produced and installed capacity data. For prediction for upcoming months, meaningful parameters will confer to, segregating data on the basis of fuel used. As we will be mainly focused on thermal (coal), hydro and nuclear for prediction, we will take only those power stations which these three as fuel.

Now, from the gathered data, we will classify it into testing and training data. Training data might range from years 2004-2012. The model generated will use the remaining data, now (testing data) for validating purposes. Also, the parameters involved in training the input will come from their respective fuel usage data.

For analyzing purpose, we would take all data and try to find the association between power plant stations of a particular state and convert the findings into a graphical and/or pie chart form. Then, the next step will be to combine the findings of all states to give a picture for the whole nation.

* 1. **Current Status**

There are no such tools at present which predict the generation of electricity with all the sources. There are many individual models which predict solar power and wind power depending on the installed plants theoretically. But no such system has been implemented which predicts the thermal and hydropower generation which are the main contributors to our power sector.

A prediction model based on fuzzy modeling has been presented at an international conference in 2011 which predicts the power generation from the farm wind generators. Power generation from a wind turbine was recorded and illustrations were made on the whole system.

Predicting solar power generation was modeled at the university of Massachusetts using machine learning algorithms with weather forecast as a parameter. But it was restricted to only the PV power generation. And has considered the only minimal amount of data for training.

Even though such models have been presented in different countries, no model has predicted the generation by all the sources for a nation. And in the country like India where utilization of electricity is very critical, there is high need of such a model which will predict the power generation state wise and by all the sources to help the power ministry to make better decisions.

* 1. **Challenges**

The main challenge is the collection of data. The data is spread across all over the internet and in different kind of formats. Collaborating and converting it into a useful data is a huge task.

Data mining is a very vast field which has many sub-topics and each sub-topic has its own pattern of solving a problem. Though it is a good thing for people to have that much of choice, it also raises the difficulty in deciding in which category a project will rely on.

The next challenge is to look for appropriate and also enough parameters, which are to be supplied to those algorithms to get the most accurate data. And looking for these parameters itself is altogether a different ball game.

**Chapter – 2**

**Literature Survey**

With the ever changing dynamics of every major corporation and a dire need to look for a pattern in that inevitable change, it was no surprise to see the ideas of data mining reaching its current position in most sought out job profile. The word *forecasting* which used to be shrouded in the cloak of mystery and magic soon started complementing the complexities of *data analysis.*

With this introduction, the learners start exploring the untested water of big data and not a lot of data, their analysis, and interpretation. During the exploration we came across many important questions that need to be answered as we moved along: -

* Does our data (electric generation for various power plants) follow some kind of a pattern?
* Can we see some kind of an association and/or covariance rule between the parameters involved?
* How deep and tightly wound up are the sets of similar data (clusters)?

And the most important of the question was whether the data are time dependent or not? This one proved to be a bit trickier to answer. But still, the search for the answer to this question led us to a very famous terminology in the field – the Time Series Analysis.

**2.1. Is it or Not?**

“Time Series Analysis of Household Electric Consumption with ARIMA and ARMA models” by Pasapitch Chujai, Nittaya Kerdprasop, and Kittisak Kerdprasop considered the household electric consumption to be a Time Series Analysis work. This analysis works on major 4 factors as mentioned in the paper. We will try to co-relate these points with our basic understanding of household electric consumption along with listing them as below:

* Seasonal effect. Now the fares of household electric consumption are majorly dependent on seasons. From India’s point of view, the summers eat the most from pocket to pay for the electricity bill. The months of that season happens to be the seasonal effect which decides one of the factors of the analysis.

The same logic can be applied to the hydropower plants, whose productivity should go down during summer and should increase relatively during monsoon. But the same cannot be said very concretely about thermal power stations, for example for those who use coal. For them, the reverse might come true, as, during monsoon, the coal mines sometimes get flooded which would decrease the input for the coal power plants.

* Cyclic. A common household does not increase its family members every year. So, it can be inferred that the consumption might be relatively comparable for the same months of two different and nearby years.

But, the same cannot be said about thermal power plants. The nuclear power plants, for example, are very much molded or destroyed by the politics, the access to the radioactive material and the ever presence of the unwanted baggage of environmental hazards. But the contradictory argument can be given for renewable resources used in their own respective power plants, as nature replenish them yearly.

* Irregular effect. The definite constant in both the worlds. The irregular effect being the occurrence of celebrations, which indirectly involves a huge show of lights and extra load on many electrical appliances.

In generation part, the damage caused by the natural disasters to the power plant stations can be considered as the irregular effect.

* This is more of a long-term and is more difficult to relate a long term of a household to that of a power plant station.

**2.2 The Techniques**

**“A Comparative Analysis of Techniques for Forecasting Electricity Consumption” by P.Ozoh, S. Abd-Rahman, J. Labadin, and M. Apperley** discusses the various ways in which one can do forecasting in electricity consumption department. Here they have extended their reach by including a much larger area (a part of Serbia) to perform forecasting. With a combination of Autoregressive Integrated Moving Average (ARIMA) along with Artificial Neural Network (ANN), they tried to find out the electricity prices using different seasons. The results were compared with the actual data along with the one obtained from Modified Newton Method. The data obtained from MNM was found to be closest (2.1%).

**2.3 The Live Example**

**“Predicting Solar Generation from Weather Forecasts Using Machine Learning” by Navin Sharma, Pranshu Sharma, David Irwin and Prashant Shenoy’s** paper dealt the topic in a very different way. They took solar intensity as one parameter along with the existing one (the weather reports and similar) to use it to generate various models for those homes that have installed the solar panels for the self-usage. This paper introduced us to the use of SVM (Support Vector Machines), a popular algorithm for reducing the dimensions of attributes. It also made us understand, how much a single parameter can change the whole prediction model. They claimed that their weather metrics (with the addition of the solar intensity parameter) is 27% more accurate than the ones created from existing ones.

**“Predicting the Energy Output of Wind Farms Based on Weather Data: Important Variables and their Correlation” by Katya Vladislavleva, Tobias Friedrich, Frank Neumann and Markus Wagner** used the data for a wind farm in Australia. They used Genetic Programming, an algorithm which locates the functions that will bind the input data to that of the output. Their project was 80% accurate.

**Chapter – 3**

**Software Specification Requirements**

**3.1 High-Level Block Diagram**

Dataset

Result

Algorithm for Analysis

Storage

Programmable Data

Data extraction

Data preprocessing

**3.2 Requirements**

**3.2.1 Hardware Interface Requirements**

Some of the hardware requirements are as follows:

* + - * Processor : Any I0ntel based processor
      * RAM : Minimum of 1GB of RAM
      * Disc Space : 2 GB for Matlab

4-6 GB for a typical installation

1 GB for Datasets

**3.2.2 Software Requirements**

Some of the software requirements are as follows:

* Operating System : Windows /8/8.1/10
* Languages : R, Python2.7, HTML/CSS
* Software and IDEs : R-Studio, Python IDLE

3.3 Functional Requirements

**F1. Prediction Capability:**

Prediction is the crux of our project. Starting from the individual power plant to the whole nation, our project should be able to predict the power that is going to be generated in future years. With the parameters like availability of co, rain, the water level at reservoir system must give out the results accurately.

**F2. Analysis Capability:**

Along with the prediction, the analysis in the form of graphical format is also necessary. The system should analyze the contribution of each state and the source in the power generation. The analysis will also consider the weather conditions to analyze the power generation.

**F3. Comparing the contribution**

* The first comparison will be based on the amount of contribution made by each state every month in power generation.
* And the second will be comparison based on the amount of contribution by each source in power generation

**F4. It should be compatible with unforeseen updates of the future**

As the time goes on more and more dataset will be accumulated, and the system should be able to accompany these datasets and should generate the updated result.

**F5. User Interface**

* The system aims at providing an efficient interface for choosing between different functionalities of the project
* The web-based interface will be simple and efficient and allow users to actively interact with the system.

**3.4 Non-Functional Requirements**

NF1. Usability

* Should be simple to operate.
* Should be of simple design. Since the system mainly demands graph and numbers, the UI must be simple, no much high-level animations and UI components are required.

**NF2. Reliability**

The system should not be very far away from the actual data.

**NF3. Adaptability**

The system must be able to cope up with the new datasets and hence it should be easy to modify the code to fit the new requirements. The role of dependencies in them should be minimum.

**NF4. Performance**

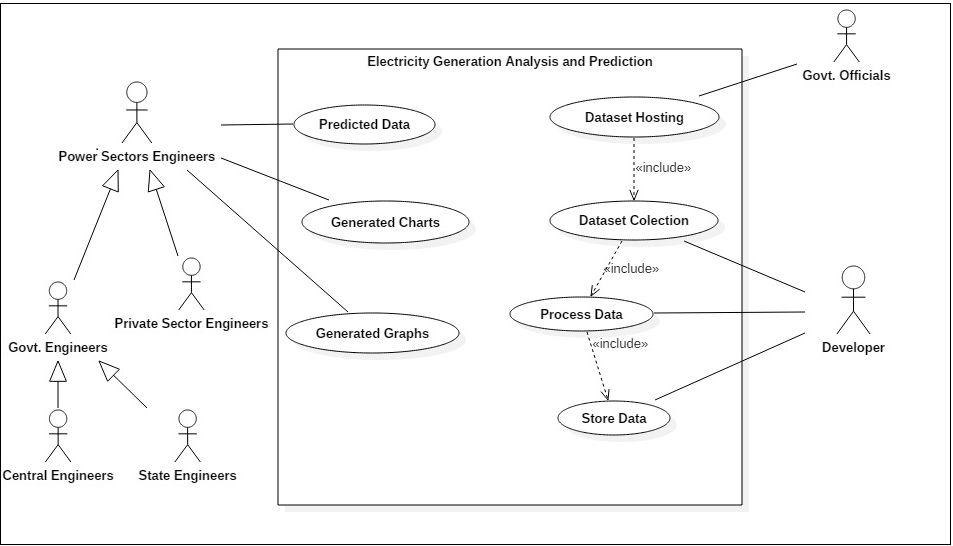
The processing time must not increase drastically after adding more data sets.

* 1. **Constraints**
* The data for the key parameters is published by the government, the minor parameters like lockouts, maintenance issue is ignored in the project.
* Natural disasters sometimes affect the plants and forces to stop the generation, which may affect the final results.
* The datasets and results should be only used by the authorized people. And should not leak the reports to any other country and should not be used for wrong purposes.
* The project is not useful for the generic purpose and common people. Although it will make a huge impact on the society, the use of the results will be only restricted to the ministry of power.
* Every plant should be working for at least 5-6 years to predict the generation of electricity by it. The new added may affect the final result.
* The contribution by diesel and gas are very less compared to thermal and hydro. And is decreasing year by year due to inefficiency and are not considered in our project.

**3.6 Assumption and Dependencies**

* The data provided by the government and other sources is assumed to be 100% authentic. Since the whole project is dependent on the datasets it is assumed to be correct.
* It is assumed that the power plants considered will be in the working condition for the following few years as the data is dependent on the present working plants.
* Although the power generation is distributed across various sectors it is under the supervision of the central government.
* The forecasting of rain and weather is considered to be accurate as the generation gets affected by these.

**3.7 Use Case Diagram**



**3.8 Testing Requirements**

* FR1/ FR2 - Prediction and analysis:

Training data might range from years 2004-2012. The model generated will use the remaining data, now (testing data) for testing purposes.

* FR3 – Comparing the Contributions

The comparisons will be done from already available datasets. The accuracy will be confirmed by manually checking generated graphs and charts.

* FR4 – Testing the Adaptability

When new datasets will be added the results might change, this will be tested manually by comparing with the old results. Thus verifying that the fluctuations are within allowed range.

* FR5 – User Interface

This requirement can be tested by confirming that the pages are displayed properly as required. Processing should not affect the interface.

**Chapter- 4**

**Schedule**