Application of Decision Tree Algorithm to study the provision of healthcare services in India: Case Study across Eight States

# ABSTRACT

This study focuses on identifying variables for decision tree regression to study the healthcare of any region and find the degree of impact each variable has on selected parameters. MMR was selected as the healthcare indicator of any region. Other parameters include - IMR, literacy rate, PHC, PNC, health expenditure, etc. Decision trees were also used to study the substitutability of parameters like institutional deliveries and skilled home deliveries. Districts with poor performance in any parameter can offset its impact by performing better in other parameters.

Keywords: Decision Tree, Linear Regression, Anemia, MMR, IMR

# Introduction

For any nation, its healthcare sector is of utmost importance, even more so for developing countries like India. However, looking at the past and current scenario, inadequacies in India’s public health sector become evident. Being the second most populous nation in the world, India faces a plethora of challenges when it comes to the provision of healthcare. Various schemes and policies have been announced and implemented over the years, but none have produced the required results to tackle these challenges. Moreover, the inequality in the services availed by the poor and the rich has been ever increasing. The recent trends indicate that in the presence of good public healthcare, people of rural and urban areas will choose it over expensive private healthcare.

India initiated the Universal Health Coverage (UHC) in the 12th Five Year Plan (2012-17) based on the recommendations of the High-Level Expert Group (HLEG). The policy ensures ‘equitable access for all Indian citizens’. It requires the provision of accessible necessary services for the population without imposing any unaffordable burden on individuals or households. Traditionally, Indian healthcare system is referred to as a mixed system where government and individuals share the burden of healthcare costs. However, the investment of the government for public provisioning of healthcare and finance social insurance in India has been extremely limited. Hence, Out-of-Pocket (OOP) expenditure by individuals is a major part of health finance.

India has often been one of the worst nations when it comes to maternal health. Maternal health is seen as the defining parameter of healthcare for any nation. A developing nation like India needs to ensure that its maternal health status is at par with the developed nations. In order to do that, it must be ensured that women are provided with adequate care. (Balarajan et al., 2011; Reddy et al., 2011)

# Objectives:

# 

1. Identification of the parameters appropriate for studying India’s healthcare system.
2. Use of various statistical tools including Linear Regression and Decision Tree Regression to assess the degree of impact of various factors on India’s healthcare.
3. This paper studies factors affecting MMR across eight states of India and shows that, irrespective of any or plural factor inefficiencies, other factors can collectively compensate and provide good results.
4. This paper identifies the relation between MMR and IMR.
5. This paper analyzes impacts of low haemoglobin levels in pregnant women.

# Literature Review:

Cost of delivering health care services in public sector, primary and community health centres in North India (Prinja et al., 2016)

The paper analyzes the cost of providing universal health care in India. The authors select three Indian states to study their primary and community health centres to find the cost of providing health care. The study found that health care delivery costs get drastically reduced if Govt. sector, PHCs and CHCs are given prime responsibility.

Inter-State Disparities in Health Care and Financial Burden on the Poor in India (Purohit, 2004)

This study aims to highlight the inter-state disparities in health care and how expenditure and policies affect a state’s health. This paper uses secondary data of over 15 major states. The study finds that the inequity in healthcare has increased over the years, and even though the urban regions are well equipped, the rural poor still remains untreated.

Inequity in hospitalization care: a study on utilization of healthcare services in West Bengal, India (Bose & Dutta, 2015)

This study focuses on the nature of in-patient care for different groups and the impact of subsidies. A benefit incidence analysis found that even though the public hospitals were extensively used, services like tests, medicines etc. could only be obtained from other expensive sources. Also, subsidies more often benefit the more well-off section of society. The out of pocket expenditure has been on the rise since independence, which has further increased inequity. The paper recommends for better provision of complementary services, in addition to the hospitalization to limit the OOP expense.

Utilization of the services of the primary health centres in India – An Empirical Study (Dar, 2015)

This study tries to find utility of the public health centres. The study was a cross-sectional research and focused on primary health centres. The study found that people in rural areas opt for PHCs over the private institutions whenever they can and a significant portion of the patients are satisfied with the treatment. The problem arises when the PHCs are ill-equipped, then people have to go the urban centres for better services. The study found that the PHCs receive only 22% of the required medicine supplies. Also, most of the centres close after 4pm and as a result during emergencies people have to go for more expensive alternatives available to them.

Quality of health care in India: Challenges, Priorities, and the road ahead (Mohanan et al., 2016)

The study finds that simply setting up health centres is in itself not enough, focus must be on improving the quality of infrastructure and services provided. The authors advocate for data driven policies with equal efforts from state and central governments.

How affordable is childbearing in India? An evaluation of maternal healthcare expenditures (Singh et al., 2016)

This paper studies the increasing cost of delivery care services faced by women. The NSSO data was used for the study. Various government schemes to ensure cheap and efficient healthcare for pregnant women were also analyzed. The study found that private hospitals often charge as high as nine times more than public hospitals. High charges are the primary reasons preventing women from seeking extensive ANC and PNC. The situation varies for every state. The findings of this paper can help study the effects of various socio-economic factors.

An Exploratory Analysis of Urban healthcare Stakeholders in India. (Chikersal, 2016)

This report provides an analytical review of the impacts on different stakeholders and their role in the nation’s mission to provide universal healthcare. The report finds that urban healthcare, unlike the rural system, is more organic and random. The non-uniformity and randomness have made the system inefficient.

Health sector reforms and changes in prevalence of untreated morbidity, choice of healthcare providers among the poor and rural population in India. (Ghosh, 2014)

This paper studies the condition of healthcare in India before and after major policy changes. The focus of this study was the rural population as the policy changes intended to bring more equity. The authors found that even for the non-poor, the financial barriers in accessing good healthcare have increased.

Inter-State Comparisons on Health Outcomes in Major States and A Framework For Resource Devolution For Health. (Rao, 2014)

The report shows that overall expenditure on healthcare is a mere 2.2% of the GPD. Even this varies across different states. States with better facilities to implement healthcare made better use of NRHM and other funds made available to them. This article suggests various changes in the present policy, including taxing medical tourism and cross-subsidizing the states.

Is Antenatal care effective in improving maternal health in rural Uttar Pradesh? Evidence from a district level household survey(Ram & Singh, 2006)

A multilevel analysis was carried to see how antenatal care affects the maternal health. Household surveys were conducted to gather the data which were further categorized according to caste, sex, literacy rate etc. The study found that women who seek antenatal care are benefited from other policies in place too. Maternal mortality risks lower significantly.

Study of impact of JSSK scheme on institutional deliveries and maternal mortality rate: Visakhapatnam district Andhra Pradesh (Revu et al., 2017)

This paper studies the impact of JSSK on maternal mortality ratio in the Visakhapatnam district. This was an observational study with data collected from 2013-17. This data was then further categorized by age, parity, educational qualification, etc. and a comparative analysis was carried out. The study found that a vast majority of women were unaware of the JSSK, and those who were aware were informed by the Anganwadi workers. MMR of less than 100 can be achieved if right steps are taken to increase awareness.

Decision Tree Methods: Applications for classification and prediction. (SONG & LU, 2015)

This paper discusses the applications of decision tree in classification and prediction problems. The paper shows how this algorithm handles the issues of variable selection and missing values. It also discusses various types of decision tree algorithms such as CART, Quest, CHAID.

# Methodology

## Parameters Used:

The following parameters were used in this study.

### Maternal Mortality Ratio:

Maternal mortality ratio refers to the number of maternal deaths every year due to various reasons, either during pregnancy, childbirth, or 42 days of childbirth per 100,000 live births per year. This is a vital indicator, which reflects the status of the health system of the nation. Maternal deaths are often due to reasons like lack of professional help, lack of nutrition, complications during pregnancy, mismanagement etc. High MMR for any region is an indication of poor healthcare facilities. (Montgomery et al., 2014)

### Infant Mortality Ratio:

Infant Mortality Ratio (IMR) refers to the number of infant deaths (under one year of age) per 1,000 live births. High IMR is a reason for concern for any region as it indicates the unavailability of basic healthcare facilities. Infant deaths can be due to various reasons like, absence of the mother, poor healthcare infrastructure, diseases, etc.

### Post-Natal Care:

Postnatal care (PNC) is the care given to the mother and her newborn baby immediately after the birth and for the first six weeks of life. Provision of proper PNC is a big issue in India. WHO recommends that mothers be provided with as much care as they were provided before the birth of the child. In India, PNC is often neglected and this leads to complications in the mothers as well as the child’s health. It is recommended that professional help must be available at all times. Research carried out in this field shows that there is a high chance that women without proper care may suffer from physical as well as mental issues. (Singh et al., 2016)

### Ante-Natal Care:

The care provided to women during the time of pregnancy by professionals is antenatal care. It can also be referred to as pregnancy care. Skilled and qualified professionals are required to carry out the process with the utmost care. Ante-Natal Care forms the core of provision of healthcare to pregnant women. This includes proper nutrition, the right medication, availability of professional help, etc. With the establishment of PHCs and CHCs India has moved a long way in the provision of ANC. (Ram & Singh, 2006)

### Literacy Rate:

A higher Literacy rate often means a greater understanding of the resources available and the desire to use them. It is often seen that illiterate women tend to neglect the services offered as unnecessary due to being ill-informed. No matter how many schemes the government launches, if the beneficiaries don’t want to benefit from these schemes, all of it becomes useless (Gokhale et al., 2002).

### Number of Institutional Deliveries:

Experts agree that the risk of stillbirth or death due to intrapartum-related complications can be reduced by about 20 percent with the presence of a skilled birth attendant. India in recent years has doubled the number of institutional deliveries, but still, a lot of scope remains. Even in the case of home deliveries the presence of skilled workers immensely improves the chance of survival of mother and child. (Gupta et al., 2012)

### Primary Health Centres:

Primary health centres are state-owned health care infrastructure and service provided by the government to ensure that poorer sections of the society are provided the basic health care facilities, free of cost. A greater number of PHCs mean better provision of healthcare in the state or district. (Dar, 2015)

### Per Capita Health Expenditure:

Higher government expenditure on health care implies more importance given to health by the government. This directly translates to better provision of healthcare to the masses. People don’t want to spend out of their pockets for health services but are willing to accept the services provided to them by the government in order to ensure their wellbeing. (Bose & Dutta, 2015)

## Obtaining the Data

In order to better understand the healthcare status of the nation, district wise healthcare data was needed. The data was obtained from various official sources present online. These include:

1. Health Management Information System, Ministry of Health (<https://nrhm-mis.nic.in/SitePages/Home.aspx>)

Extensive district-wise data of over 500 parameters were present on this portal. The 2018-19 data was used for analysis purposes. The data comprised of parameters like ANC participation, PNC participation, MMR, IMR, JSSK, Haemoglobin levels etc.

1. Census India (<http://censusindia.gov.in/>)

District wise population and literacy rate data were obtained from this portal.

1. EPWRF India Time Series (<http://www.epwrfits.in/>)

State-wise healthcare expenditure data was obtained from this portal.

## Cleaning the Data:

The data obtained was disaggregated and from different sources. It then had to be assembled and cleaned. This was done using VBA and Python. The script used has been attached in the appendix A and B.

Table 1: Variable names and their description

|  |  |  |
| --- | --- | --- |
| S.no. | Variable Name | Variable Description |
| 1 | State | Name of State |
| 2 | District | Name of district |
| 3 | Population | Population of the district |
| 4 | Calcu mmr | Maternal mortality rate of the district (per 100, 000 live births) |
| 5 | Calcu IMR 12 | Infant mortality rate of the district (per 1,000 live births) |
| 6 | Pnc per live birth | Number of women receiving Postnatal care per live birth |
| 7 | Per capita health exp | Per capita health expenditure of the state |
| 8 | Phc per 100000 | Number of primary health centres in the district per 100,000 |
| 9 | literacy | Literacy rate of the district |
| 10 | Per home skill | Percent of home deliveries attended by skilled workers |
| 11 | Percent insti | Percent of total deliveries carried out in a healthcare institution |
| 12 | haemo | Percent of pregnant women registered for ANC with haemoglobin levels less than 7g/ml |

Eight states were chosen for analysis purpose. These aer:

* Andhra Pradesh
* Goa
* Haryana
* Karnataka
* Kerala
* Madhya Pradesh
* Maharashtra
* Odisha

## Analysis of the Data:

Python was used for data analysis. The data description after cleaning has been attached in the appendix C.

The next step was to handle missing values. Missing values for any district were replaced with the mean of the values belonging to the same state. Furthermore, the parameters were of different units and the parameters were significantly skewed. Parameters with absolute skewness greater than 0.75 were normalized using box-cox transformation. Following this, the data was ready for analysis.

The correlation matrix was first plotted to study the correlation between the variables, which gave an idea of how strongly the variables were related. Scatter plots for all the variable relationships were then plotted to get a clear picture of the variable relationships. The plots are attached shown below. Having an idea of how variables interacted with each other gives a platform for greater in-depth analysis.

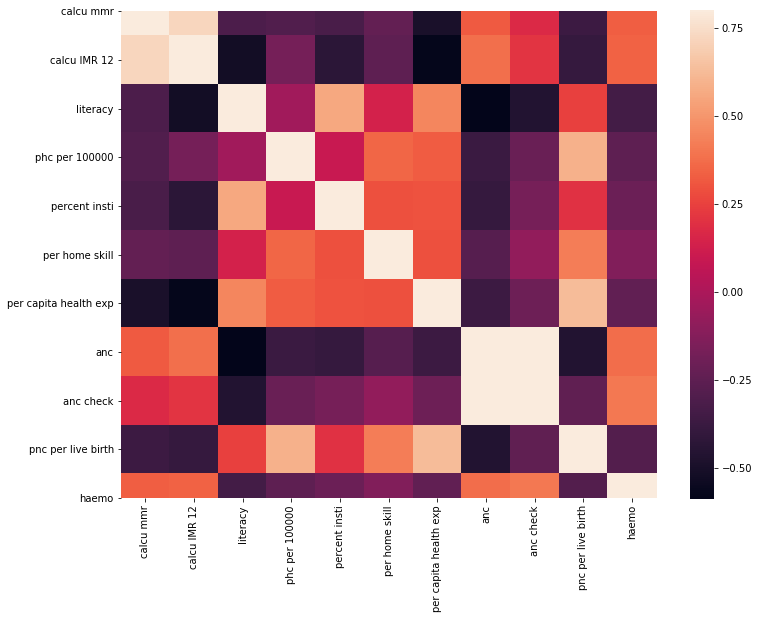


Figure 1: Correlation Matrix

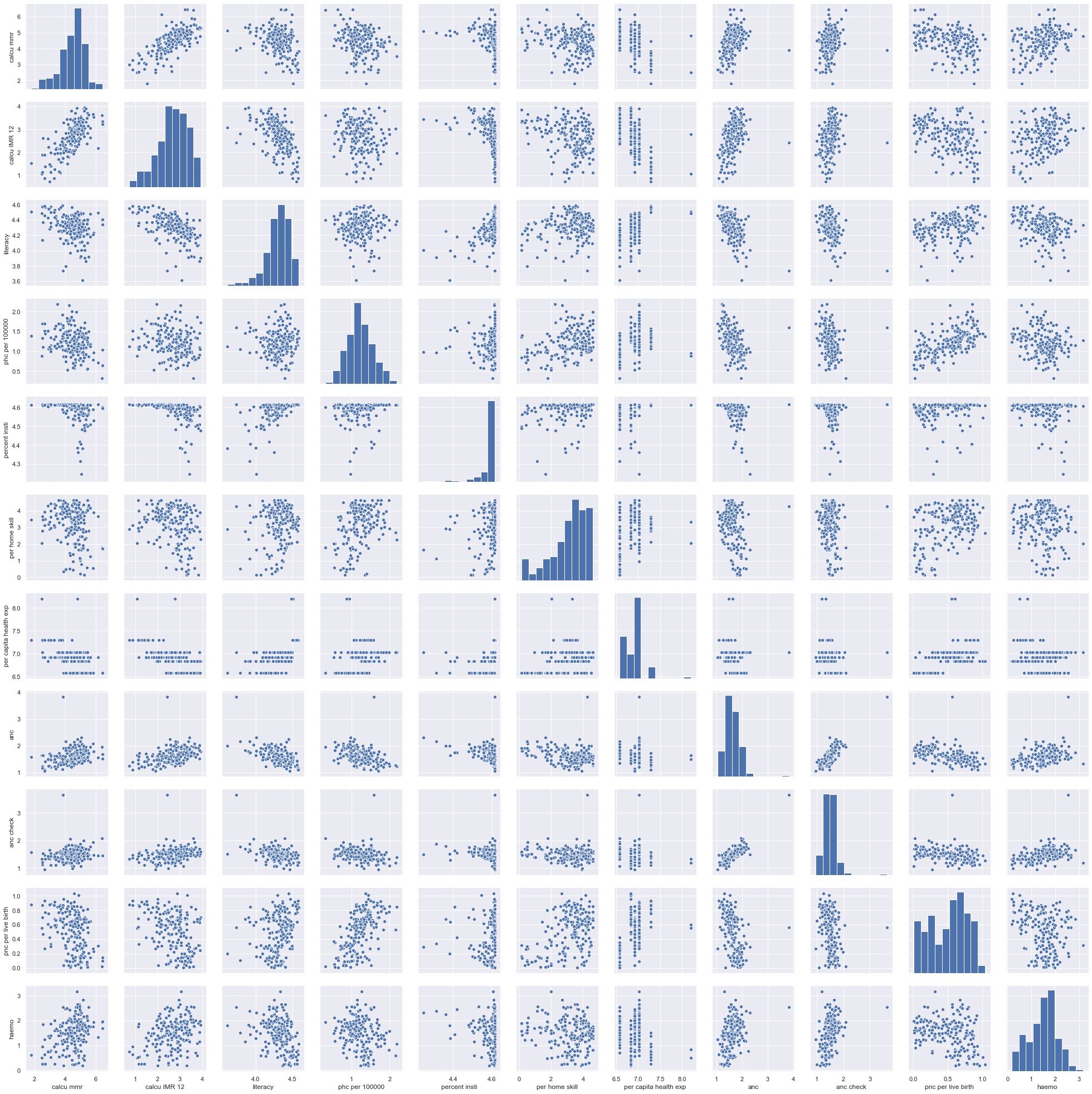
[](https://drive.google.com/file/d/1NizS7_RmvgiuXuIuJI9tRC3QhgVJpgmX/view?usp=sharing)

Figure 2: Combined Scatter Plot

For further analysis, the algorithms used were:

* Linear Regression
* Decision Tree Regression

### Linear Regression:

Linear regression is a supervised learning algorithm where a linear relationship between two or more variables is modeled. In this study linear regression is used to identify the linear relationships between variables like MMR vs IMR.

### Decision Tree Regression:

Decision tree is a type of supervised learning algorithm that can be used in both regression and classification problems. It works for both categorical and continuous input and output variables. Decision Tree is a flow-chart like structure, where the leaf nodes show the final result, and the remaining nodes are used to split the tree based on the condition specified in the node. Decision Trees in this study give the expected MMR for particular values of parameters and help in identifying the substitutability of parameters.

The decision tree looks like the picture given below.

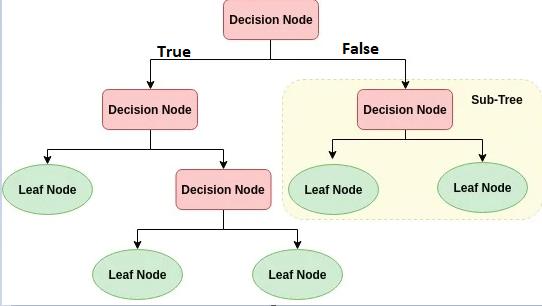


Figure 3: Decision Tree Representation

The decision node looks like the picture given below.

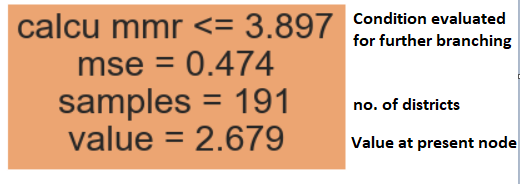


Figure 4: Representation of a node in the tree

## Analysis Results:

### Result Sheet 1:

The first decision tree was plotted to analyze how different parameters affected MMR in the district. The following parameters were used:

1. Literacy rate
2. Percent of institutional deliveries out of total deliveries
3. Percent of home deliveries attended by skilled workers
4. Per Capita Health Expenditure in State.

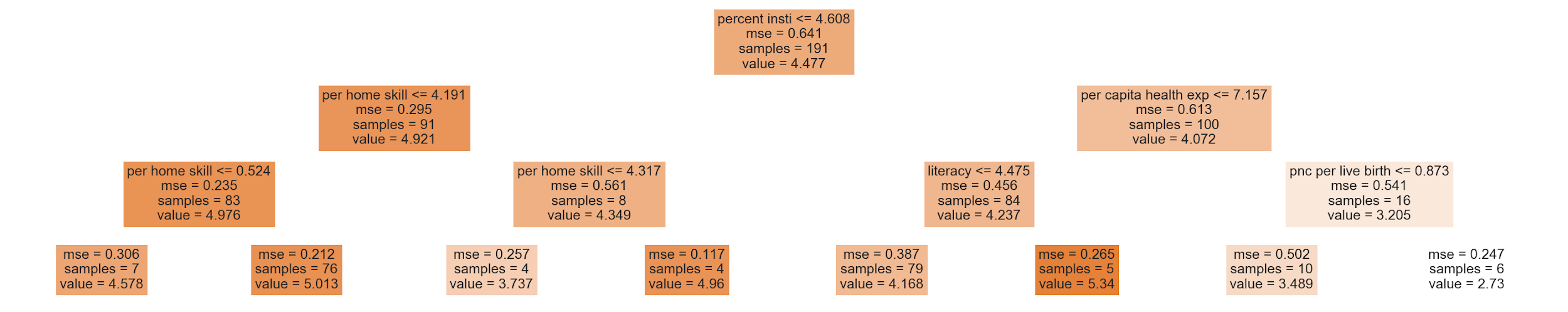
[](https://drive.google.com/file/d/1ggxekfHTANKDX02T4bSjLyNJcYq3kLP1/view?usp=sharing)

Figure 4: Decision Tree 1 studying the impact of four parameters on MMR

The tree reveals a great deal about how the parameters affect the MMR, and gives an idea of how parameters can be substituted for the desired results. The root node splits on the basis of percent institutional deliveries, indicating its strong impact on MMR. This is further verified in the scatter plot of ‘Percent Institutional deliveries’ and ‘MMR’.

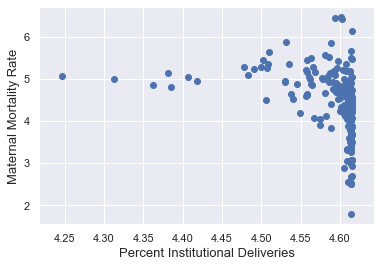


Figure 5: Scatter Plot 1 (MMR vs percent insti deliveries)

The split at the root node is shown in the table below.

Table 2: Decision Tree 1 - the impact of institutional deliveries

|  |  |  |
| --- | --- | --- |
| Percent Institutional Deliveries | Expected MMR | Samples |
| <= 4.608 | 4.921 | 91 (out of 191) |
| > 4.608 | 4.072 | 100 (out of 191) |

**Interpretation:**

Districts with ‘Percent Institutional Deliveries (PID)’ less than 4.608 have an expected MMR of 4.921, whereas districts with PID greater than 4.608 have an expected MMR of 4.072. This is in line with the scatter plot.

Similarly, another important observation is the impact of ‘Per Capita Health Exp.’ on MMR. The scatter plot verifies the negative relationship between ‘Per Capita Health Exp’ and MMR.

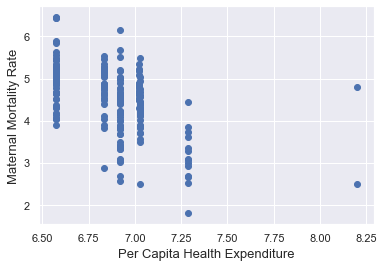


Figure 6: Scatter Plot for MMR vs Per capita health exp

Table 3: Decision Tree 1 - impact of Per capita health exp

|  |  |  |
| --- | --- | --- |
| Per Capita Health Exp  ( Percent Institutional Deliveries >4.608 ) | Expected MMR | Samples |
| <= 7.157 | 4.237 | 84 (out of 100) |
| > 7.157 | 3.205 | 16 (out of 100) |

**Interpretation:**

*(This table is for the branch of the tree with ‘Percent Institutional Deliveries’ > 4.608.)*

Districts with ‘Per Capita Health Exp’ less than 7.157 have an expected MMR of 4.237, whereas districts with ‘Per Capita Health Exp’, greater than 7.157 have an expected MMR of 3.205.

The decision tree plotted in this analysis (Result Sheet 1) shows how different parameters like literacy rate, Health Expenditure, etc. impact the MMR. It further highlights how one parameter can offset the absence of another parameter. The parameter substitution is further studied in greater detail in the next series of results.

### Result Sheet 2:

The analysis performed in ‘Result Sheet 1’ shows poor performance in any parameter can be offset by other parameters. In the following analysis the substitutability of ‘Per Capita Health Exp’ and ‘PNC per live birth’ is studied. Decision Tree Regression is performed with ‘Per Capita Health Exp’ and ‘PNC per live birth’ as explanatory variables.

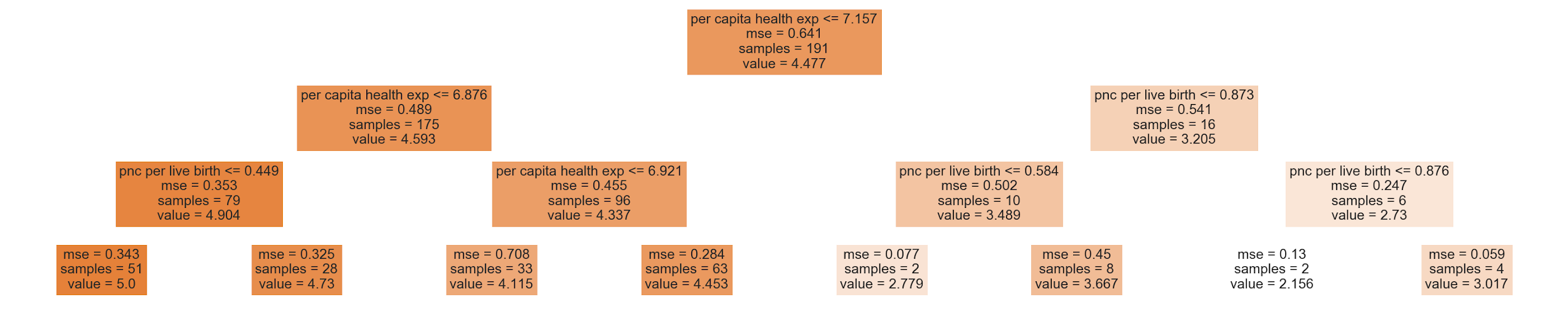
[](https://drive.google.com/file/d/18O95txM-n5XST3VS9MUKlxarPc-Mo6Ga/view?usp=sharing)

Figure 7: Decision Tree 2 - Substitutability of PNC and Per capita health exp

Table 4: Decision Tree 2 - Substitutability of PNC and Per capita health exp

|  |  |  |  |
| --- | --- | --- | --- |
| Per Capita Health Exp | PNC per live birth | Expected Normalized MMR | Number of districts |
| > 7.157 | - | 3.205 | 16 |
| <= 7.157 | - | 4.593 | 175 |
| <= 6.876 | - | 4.904 | 79 |
| Between 6.876 and 7.157 | - | 4.337 | 96 |
| <= 6.876 | <= 0.449 | 5.0 | 51 |
| <= 6.876 | > 0.449 | 4.73 | 28 |

**Interpretation:**

The tree at the root node branches on the basis of ‘Per Capita Health Exp’. Districts with ‘Per Capita Health Exp’ > 7.157 have an expected MMR of 3.205, whereas districts with ‘Per Capita Health Exp’ <= 7.157 have an expected MMR of 4.593. The left branch is then further split on the basis of ‘Per Capita Health Exp’. Districts with ‘Per Capita Health Exp’ > 6.876 have an expected MMR of 4.337, whereas districts with ‘Per Capita Health Exp’ <= 6.876 have an expected MMR of 4.904. As expected the expected MMR increases with decreasing ‘per capita health exp’. The fourth node sees branching on the basis of ‘PNC per live birth’ for districts with ‘per capita health exp’, <= 6.876. Districts with ‘PNC per live birth’ <= 0.449 have an expected MMR of 5.0, whereas districts with ‘PNC per live birth’ > 0.449 have an expected MMR of 4.73.

‘Per capita health exp’ is an important parameter that directly impacts the expected MMR, but states which can’t afford higher ‘per capita health exp’ should try to improve the PNC facilities and educate people about benefits of good post-natal care. The availability of good PNC care can offset less ‘per capita health exp’.

Similarly, the following decision tree shows the substitutability of ‘Institutional deliveries’ with ‘Home Deliveries in presence of skilled help’.

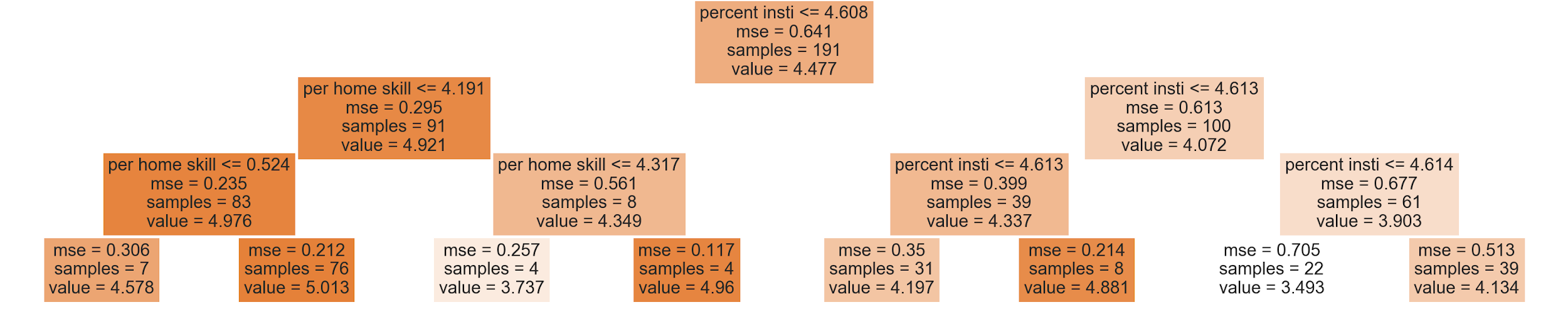
[](https://drive.google.com/file/d/1DJ9Dul37e5iJJaVXFxCYXchdZtGQYQLT/view?usp=sharing)

Figure 8: Decision Tree 3 - Institutional deliveries vs skilled home deliveries

Table 5: Decision Tree 3 - Institutional deliveries vs skilled home deliveries

|  |  |  |  |
| --- | --- | --- | --- |
| Percent Institutional Deliveries | Percent deliveries at home by skilled help | Expected MMR | Number of districts |
| > 4.608 | - | 4.072 | 100 (out of 191) |
| <= 4.608 | - | 4.921 | 91 (out of 191) |
| <= 4.608 | <= 4.191 | 4.976 | 83 (out of 91) |
| <= 4.608 | > 4.191 | 4.349 | 8 (out of 91) |

**Interpretation:**

The tree at the root node branches on the basis of ‘Percent Institutional Deliveries’. Districts with ‘Percent Institutional Deliveries’ > 4.608 have an expected MMR of 4.072, whereas districts with ‘Percent Institutional Deliveries’ <= 4.608 have an expected MMR of 4.921. The left branch of the tree then splits on the basis of ‘Percent deliveries at home by skilled help’. Districts with ‘Percent deliveries at home by skilled help’ <= 4.191 have an expected MMR of 4.976, whereas districts with ‘Percent deliveries at home by skilled help’ > 4.191 have an expected MMR of 4.349.

It can be observed that the absence of institutional deliveries can be offset by having skilled help during home deliveries. Healthcare institutions provide infrastructure and good services to pregnant women, leading to lower MMR. But not all regions can ensure institutional deliveries to all women, due to various reasons. In such cases ensuring the availability of skilled help for home deliveries can help pregnant women.

### Result Sheet 3:

IMR is a good indicator of the health status of any region. It is of utmost importance for any society to ensure proper health and wellbeing of newborns. The scatter plots indicated that IMR was directly correlated with MMR, literacy rate, Post Natal Care.

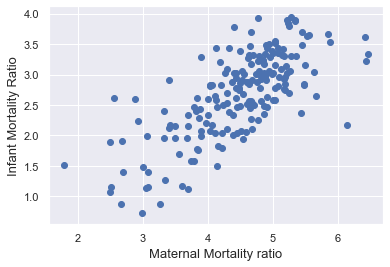


Figure 9: Scatter plot of IMR vs MMR

Regression analysis of MMR and IMR gives the following results:

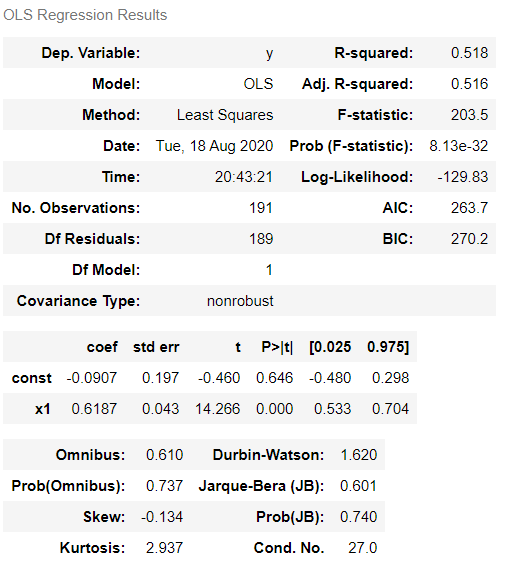


Figure 10: Summary of Regression analysis of IMR vs MMR

R-squared of 0.516 and confidence interval of 0.533 to 0.704, shows the strong relation of the independent variable.

The intercept is -0.0907 and the coefficient is 0.6187.

IMR = -0.0907 + ( 0.6187 \* MMR ).

The presence of the mother directly benefits an infant, ensuring his / her wellbeing.

Decision Tree Regression analysis of IMR with parameters MMR and Post-Natal Care is shown below.

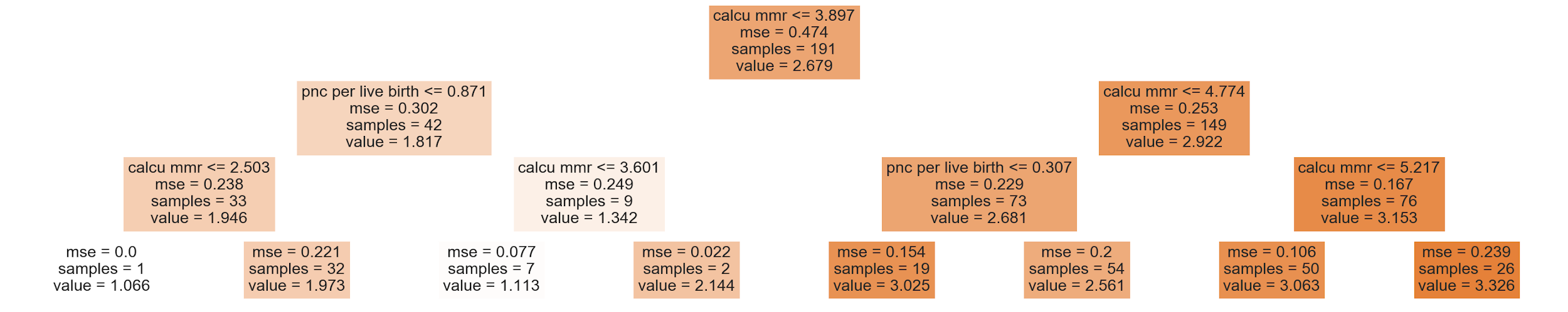
[](https://drive.google.com/file/d/1KoYp7bvfPGz1ugQ7HJdsB2lv_z2C6zrp/view?usp=sharing)

Figure 11: Decision Tree 4 - Impact of MMR and PNC on IMR

Table 6: Decision Tree 4 - Impact of MMR and PNC on IMR

|  |  |  |  |
| --- | --- | --- | --- |
| MMR | PNC per live birth | Expected IMR | Number of districts |
| <= 3.897 | - | 1.817 | 42 (out of 191) |
| Between 3.897 and 4.774 | - | 2.681 | 73 (out of 149) |
| Between 3.897 and 4.774 | <= 0.307 | 3.025 | 19 |
| Between 3.897 and 4.774 | > 0.307 | 2.561 | 54 |

**Interpretation:**

Districts with MMR less than 3.897 have an expected IMR of 1.817, whereas districts with MMR between 3.897 and 4.774 have an IMR of 2.681. Districts with higher MMR have high IMR, but this can be offset with high PNC. Districts with high MMR along with PNC greater than 3.07 have lower IMR at 2.561.

### Result Sheet 4:

Haemoglobin levels of women is often used as an indicator to get an overview of the health status of women in society. Haemoglobin level less than 7gm/dl indicates severe anaemia.

Linear Regression analysis and corresponding scatter plot shows the correlation between MMR and low haemoglobin levels.

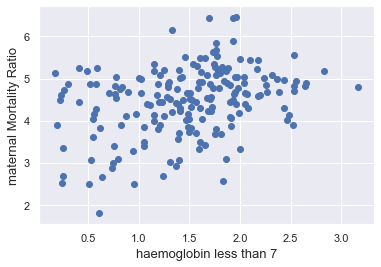


Figure 12: Scatter Plot of MMR vs Low-Haemoglobin levels

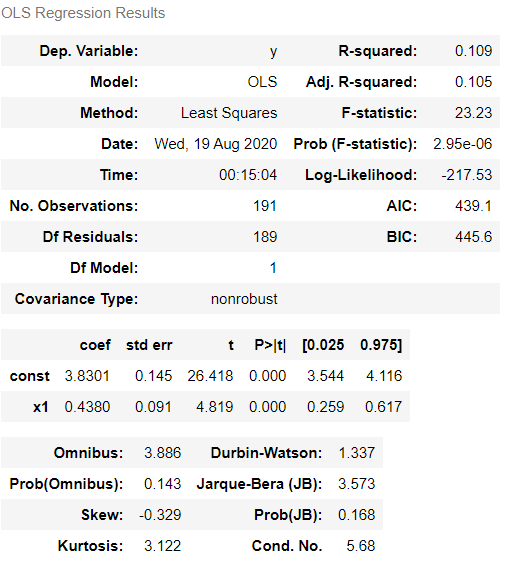


Figure 13: Regression summary of MMR vs Low-haemoglobin levels

R-squared of 0.109 and confidence interval of 0.259 to 0.617, shows the strong relation of the independent variable.

The intercept is 3.8310 and the coefficient is 0.4380.

MMR = 3.8310 + ( 0.4380 \* haemo ).

Haemoglobin levels in women depend on a lot of factors like the literacy rate, availability of Primary Health Centres, etc. The decision tree plotted with these variables gave the following output.

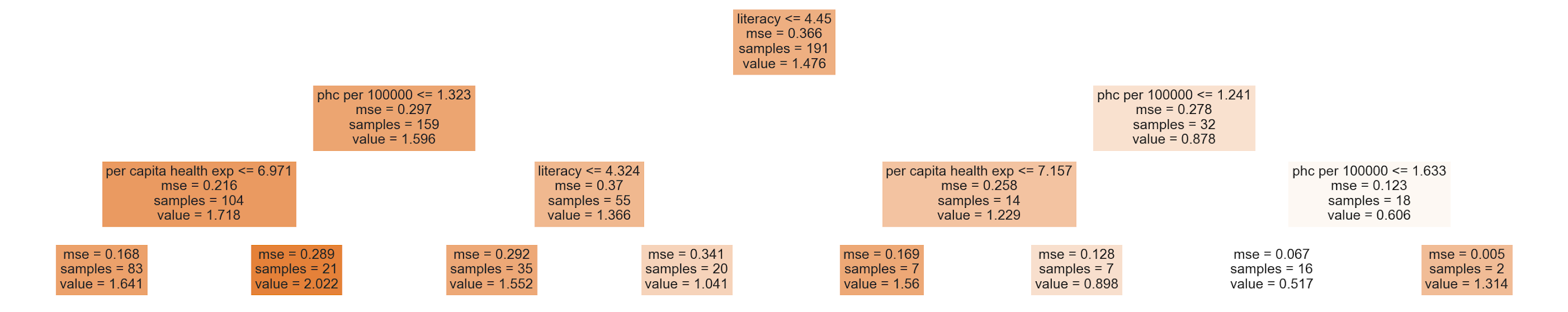
[](https://drive.google.com/file/d/132ftOoFc5PZY8fn8tqUqi7dyJJOehXOQ/view?usp=sharing)

Figure 14: Decision Tree 5 - impact of literacy rate and PHC on low-haemoglobin rates

Table 7: Impact of literacy rate and PHC on low-haemoglobin rates

|  |  |  |  |
| --- | --- | --- | --- |
| Literacy Rate | PHC per 100000 | Percent Women with Haemoglobin less than 7 (anemia) | Number of districts |
| > 4.45 | - | 0.878 | 32 (out of 191) |
| <= 4.45 | - | 1.596 | 159 (out of 191) |
| <= 4.45 | <= 1.323 | 1.718 | 104 (out of 159) |
| <= 4.45 | > 1.323 | 1.366 | 55 (out of 159) |

**Interpretation:**

Districts with literacy rates higher than 4.45 have anemia rate of 0.878, whereas districts with literacy rates less than 4.45 have anemia rate of 1.596. Lower literacy rate levels keep women ill-informed about anemia, resulting in poor health choices. It can be observed from the decision tree that districts with lower literacy rates can still help women by ensuring the availability of PHCs. Districts with literacy rate less than 4.45 but ‘PHC per 100,000’ greater than 1.323 have anemia rates of 1.366.

# Result Analysis:

This study was done on district level to identify the health status of the regions. Maternal health was chosen as the parameter to check the health status of the regions. The use of machine learning techniques like Linear Regression and Decision Tree Regression made the analysis easier to process and understand.

The Correlation matrix and the combined scatter plots gave an idea of how the parameters interact with each other. Decision tree was then used for analysis, which showed that higher per capita health expenditure and institutional deliveries can lower the MMR. The second analysis focused on the substitutability of the parameters. The decision trees plotted in this analysis showed that despite fewer institutional deliveries low MMR is still possible by ensuring high skilled home deliveries. Similarly, low per capita health expenditure can be offset by providing good PNC to women.

The third result sheet deals with the relation between MMR and IMR. The regression analysis performed shows that an IMR is directly correlated with MMR. This further highlights the importance of maternal health. A newborn child’s survival is to a great extent dependent on the mother’s survival.

The fourth result sheet uses regression analysis to show how high anemia rates are correlated with high MMR. Decision Tree Regression was then used, which showed districts with high literacy rates and the availability of PHC have low anemia rates.

# References

Balarajan, Y., Selvaraj, S., & Subramanian, S. (2011). Health care and equity in India. *The Lancet*, *377*(9764), 505–515. https://doi.org/10.1016/S0140-6736(10)61894-6<https://doi.org/10.11919/j.issn.1002-0829.215044>

Bose, M., & Dutta, A. (2015). Inequity in hospitalization care: A study on utilization of healthcare services in West Bengal, India. *International Journal of Health Policy and Management*, *4*(1), 29. https://doi.org/10.15171/ijhpm.2015.05

Chikersal, A. (2016). *An Exploratory Analysis of Urban Healthcare Stakeholders in India*.

Dar, K. H. (2015). *Utilization of the Services of the Primary Health Centres in India—An Empirical Study*. 5.

Ghosh, S. (2014). Health sector reforms and changes in prevalence of untreated morbidity, choice of healthcare providers among the poor and rural population in India. *International Journal of Health Policy and Management*, *2*(3), 125–130. https://doi.org/10.15171/ijhpm.2014.32

Gokhale, M. K., Rao, S. S., & Garole, V. R. (2002). Infant Mortality in India: Use of Maternal and Child Health Services in Relation to Literacy Status. *Journal of Health, Population and Nutrition*, *20*(2), 138–147. JSTOR.

Gupta, S. K., Pal, D. K., Tiwari, R., Garg, R., Shrivastava, A. K., Sarawagi, R., Patil, R., Agarwal, L., Gupta, P., & Lahariya, C. (2012). Impact of Janani Suraksha Yojana on Institutional Delivery Rate and Maternal Morbidity and Mortality: An Observational Study in India. *Journal of Health, Population, and Nutrition*, *30*(4), 464–471.

Mohanan, M., Hay, K., & Mor, N. (2016). Quality Of Health Care In India: Challenges, Priorities, And The Road Ahead. *Health Affairs*, *35*(10), 1753–1758. https://doi.org/10.1377/hlthaff.2016.0676

Montgomery, A., Ram, U., Kumar, R., & Jha, P. (2014). Maternal Mortality in India: Causes and Healthcare Service Use Based on Nationally Representative Survey. *PloS One*, *9*, e83331. https://doi.org/10.1371/journal.pone.0083331

Prinja, S., Gupta, A., Verma, R., Bahuguna, P., Kumar, D., Kaur, M., & Kumar, R. (2016). Cost of Delivering Health Care Services in Public Sector Primary and Community Health Centres in North India. *PLOS ONE*, *11*(8), e0160986. https://doi.org/10.1371/journal.pone.0160986

Purohit, B. C. (2004). Inter-State Disparities in Health Care and Financial Burden on the Poor in India. *Journal of Health & Social Policy*, *18*(3), 37–60. https://doi.org/10.1300/J045v18n03\_03

Ram, F., & Singh, A. (2006). IS ANTENATAL CARE EFFECTIVE IN IMPROVING MATERNAL HEALTH IN RURAL UTTAR PRADESH? EVIDENCE FROM A DISTRICT LEVEL HOUSEHOLD SURVEY. *Journal of Biosocial Science*, *38*(04), 433. https://doi.org/10.1017/S0021932005026453

Rao, S. (2014). *Inter-State Comparisons on Health Outcomes in Major States and a Framework for Resource Devolution for Health.*

Reddy, K. S., Patel, V., Jha, P., Paul, V. K., Kumar, A. S., & Dandona, L. (2011). Towards achievement of universal health care in India by 2020: A call to action. *The Lancet*, *377*(9767), 760–768. https://doi.org/10.1016/S0140-6736(10)61960-5

Revu, S., Kodali, V. R., & Dongabanti, H. D. (2017). Study of impact of JSSK scheme on institutional deliveries and maternal mortality rate: Visakhapatnam district Andhra Pradesh. *International Journal of Research in Medical Sciences*, *6*(1), 160. https://doi.org/10.18203/2320-6012.ijrms20175713

Singh, P., Kumar, V., & Verma, S. (2016). How Affordable is Childbearing in India? An Evaluation of Maternal Healthcare Expenditures. *Newborn and Infant Nursing Reviews*, *16*(4), 175–183. https://doi.org/10.1053/j.nainr.2016.09.012

SONG, Y., & LU, Y. (2015). Decision tree methods: Applications for classification and prediction. *Shanghai Archives of Psychiatry*, *27*(2), 130–135. https://doi.org/10.11919/j.issn.1002-0829.215044

# Appendix A

**VBA Script**

Sub GetSheets()

Path = ""

Filename = Dir(Path & "\*.xls")

Do While Filename <> ""

Workbooks.Open Filename:=Path & Filename, ReadOnly:=True

For Each Sheet In ActiveWorkbook.Sheets

Sheet.Copy After:=ThisWorkbook.Sheets(1)

Next Sheet

Workbooks(Filename).Close

Filename = Dir()

Loop

End Sub

# Appendix B

**Python script**

# -\*- coding: utf-8 -\*-

"""

Created on Fri Aug 7 16:40:46 2020

@author: AbhishekKumar

"""

import os

import pandas as pd

path = os.getcwd()

#path = "E:\\BITS\\sem 8\\2018-2019\\Arunachal Pradesh\\"

#os.chdir("E:\\BITS\\sem 8\\2018-2019\\Arunachal Pradesh\\")

files = os.listdir(path)

files

data = pd.read\_excel('dadra and nagar haveli.xlsx', 'Sheet1 (2)')

final = data.iloc[:, [2, 54]]

final = final.transpose()

for f in files:

xls=pd.ExcelFile(f)

res = len(xls.sheet\_names)

for i in range(res):

sheet = xls.parse(i)

data = sheet.iloc[:, [54]]

data = data.transpose()

frames = [final, data]

final = pd.concat(frames)

for f in files:

xls = pd.ExcelFile(f)

res = len(xls.sheet\_names)

for i in range(res):

count+=1

sheet = xls.parse(i, header=None)

mystring = sheet.iloc[0][0]

keyword = '-> '

before\_keyword, keyword, after\_keyword = mystring.partition(keyword)

keywordd = '- '

before\_keywordd, keywordd, after\_keywordd = before\_keyword.partition(keywordd)

temp = [after\_keywordd, after\_keyword]

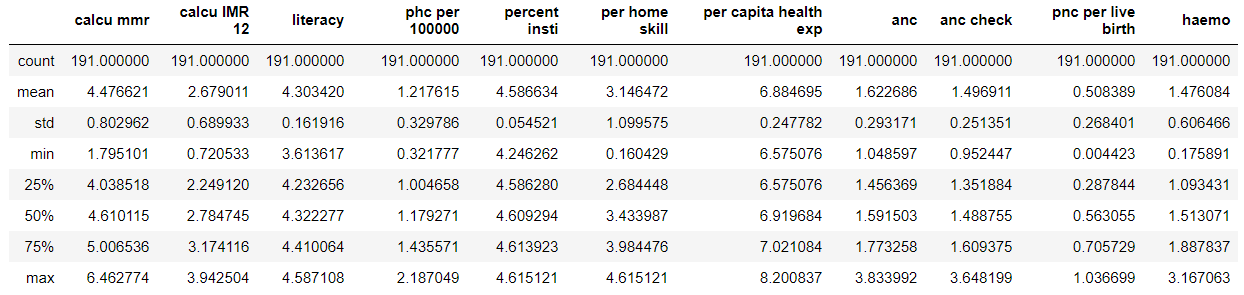
tempdf = pd.DataFrame([temp])

frames = [df, tempdf]

df = pd.concat(frames)

# Appendix C

**Data Description**



# Appendix D

**Analysis using Python**

<https://drive.google.com/file/d/11TG7mlLD2Wa2_GUc6qk0DCmwqtqTMohc/view?usp=sharing>