Modeling Health Insurance Selection in Indian Market using Data Mining Approach
Pratik Biswas¹ and Partha Sarathi Bishnu²

Health insurance costs across the world have increased alarmingly in recent years, just as the number of insurance companies and their health care schemes have also gone up. In view of high expense in medication, health insurance policy is very essential for individuals today. These policies are provided by both insurance companies and banking sectors with different medical plans. The selection of appropriate medical insurance plan by a customer has been becoming a difficult task day by day. Perhaps this is the reason why a website like policybazaar.com is so popular from customer’s point of view. In this paper, authors made an attempt to model health insurance policy selection using Data mining approach.

Keywords: Data mining, Customer satisfaction, Healthcare.

Introduction
Data mining can be defined as a process that uses a variety of data analysis tools to discover patterns and relationships in data that may be used to make valid predictions. Data mining tools are based on highly automated search procedures. In present era various insurance companies are producing enormous amounts of information regarding medical facilities provided by them which are difficult to select. So, there is a need of powerful automated data mining tools for analysis and interpreting the useful information from this data. This information is very valuable for a customer to understand the cause of diseases and the benefits provided by the companies in order to cost effective treatment to patients. Data mining offers novel information regarding healthcare which in turn helpful for making financial as well as medical decision. We describe our framework, problem formulation, evaluation metrics, and experimental results on the basis of data from a large range of health insurer provided by different banking sectors like Bank of India, Punjab national bank, State bank of India with collaboration with national insurance, the oriental insurance company limited.

Because of the rapid progress of information technology, the amount of information stored in insurance databases is rapidly increasing. These huge databases contain a wealth of data and constitute a potential goldmine of valuable business information. Health insurance becomes a major field for insurance companies now a days. Different companies are providing different lucrative scheme to fetch the customer’s attention. These schemes are full with medical benefits. This article addresses the selection of the health insurance plan in customer point of view. We first propose a table to gather various medical services (facilities) provided by different banks and insurance companies. Then, with this table, we applied data mining technique to extract best medical information for the customer. The outcome shows several interesting results, which suggest that the reuse of stored data will give a powerful tool to improve the decision quality taken by the customer.

¹ Faculty Member, IIT, Kharagpur, E-mail ID: Pratik_biswas2007@rediffmail.com
² Assistant professor, BIT, Ranchi, E-mail ID: pubishnu@gmail.com
Why Health Insurance Is Important?
As healthcare is becoming very costly now days, having health insurance is important because medical coverage helps people get timely medical care and improve their lives and health. It is also found that uninsured people receive inadequate medical care and also not in time, risking their health. Lack of medical insurance is a fiscal burden for uninsured people and their families. The benefits of expanding coverage outweigh the costs for added services.

1.2 Information about Medical Plans Proposed by Insurance Companies
We are data rich, but information poor. Cashless Bima policy is an unique Health insurance Policy designed especially for the account holders of different Banks of India [1][2]. The entire family consisting of the account holder, spouse and two dependent children can be covered under this policy. It generally covers the following services: Incentive care unit expenses. Room stay including I.C.U, Surgeon, Anesthetist, Medical Practitioner, Consultants, Specialists Fees, Anesthesia, Blood, Oxygen, Operation Theatre Charges, Surgical Appliances, Medicines & Drugs, Dialysis, Chemotherapy, Radiotherapy, Artificial Limbs, Cost of Prosthetic devices implanted. Reimbursement of Funeral Expenses, in case of death of the insured person during surgical procedure like pacemaker, Relevant Laboratory/Diagnostic test, X-Ray etc [1]. Ambulance services In general 1% of the sum insured. Hospital cash, Reimbursement of incidental expenses during the period of hospitalization during the policy period to an insured peril covered under the period. Domiciliary Hospitalization: Surgeon, Medical Practitioner, Consultants, Specialists fees, blood, Oxygen, Surgical Appliances, Medicines & Drugs, Diagnostic Material and Dialysis, Chemotherapy, Nursing expenses during policy period [1][2]. A fixed amount of premium during the benefit period, an insured person pays before the insurer starts to make payments for covered medical services. Plans may have both per individual and family deductibles. Flexible benefit plans are also available which offer employees a choice between permissible taxable benefits, including cash, and nontaxable benefits such as life and health insurance, vacations, retirement plans and child care [1][2].

1.3 Tree-Based Model

![Diagram]

**Figure1.** Customer attribute on the basis of socio economical status
In figure 1 customer's attribute is shown on the basis of socio economical status, here customers are broadly divided into four categories, Salary: High income group. Middle income group and lower income group. Age: below 17, years in between 18 to 64 years and senior citizen that is above 65. Sex: whether male or female and at last dependent or not. It is made because health insurance policies are also following certain rules. A customer can have the maximum benefit only when the match between the policy and requirement is perfect.

2. Methodology

We now discuss about the selection of medical insurance by applying data mining technique.

2.1. Data acquisition: The first step is to gather information about the benefits given by the various insurance companies either by verbal communication or from webpage of the company. We construct an initial table, “Table 1”, where summary information regarding policies and diseases are plotted with the following conditions [12]:

\[ P_i \cdot D_j \begin{cases} = 1, & \text{when policy } P_i \text{ covers disease } D_j \\ = 0, & \text{policy } P_i \text{ does not cover disease } D_j \end{cases} \]

For, i = 1, 2, ..., n and j = 1, 2, ..., m.

n is no. of policies and m is no. of diseases.

2.2. Customer Preference Data: Customers are divided in to four groups on the basis of their socio economical status, and then the data (information) are processed on the basis of their priorities. Here we construct another table on the basis of distance calculation (d_k) [k = 1, 2, ..., n] between policies and customer preference using formula [12]:

\[ d_{ij} = \sqrt{(x_{i1} - x_{j1})^2 + (x_{i2} - x_{j2})^2 + \ldots + (x_{in} - x_{jm})^2} \]

Where \( i = (x_{i1}, x_{i2}, \ldots, x_{in}) \) and \( j = (x_{j1}, x_{j2}, \ldots, x_{jm}) \) are two n dimensional data object.

We had selected \( \min(d_k) \) for \( k = 1, 2, \ldots, n \), the \( k^{th} \) policies will be best selection for the customer.

Table 2 indicates customer's preferences towards coverage of diseases by "1" and dissatisfaction by "0". Here selection is done by one particular set of customers, which may change according to the change in priority of the customer.

2.3 Selection Method:

In figure 2 collection of information and their comparison is shown, next we select the best one on the basis of the distance formula using equation 2 and the minimum distance indicates the customer's preference and plotted in table 2.

3. Conclusion:

It may be noticed that there are anomalies in the selection of policies since expectations of different types of customers from health insurance are different. For some, it may be only healthcare benefits whereas for some it is tax benefit that is more important. It is found that whether it is a private insurance company or a government organization, facilities given are similar, with the only difference being in their terms and conditions and premium amount. However, each person would like to take only one policy. Hence, the suggested technique is useful for appropriate policy selections by different profiles of customers.
Table 1: Policies against diseases database.

<table>
<thead>
<tr>
<th>$D_1$</th>
<th>$D_2$</th>
<th>$D_3$</th>
<th>$D_4$</th>
<th>$D_5$</th>
<th>$D_6$</th>
<th>$D_7$</th>
<th>$D_8$</th>
<th>$D_9$</th>
<th>$D_{10}$</th>
<th>$D_{11}$</th>
<th>$D_{12}$</th>
<th>$D_{13}$</th>
<th>$D_{14}$</th>
<th>$D_{15}$</th>
<th>$D_{16}$</th>
<th>$D_{17}$</th>
<th>$D_{18}$</th>
<th>$D_{19}$</th>
<th>$D_{20}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2: Customer preference table

<table>
<thead>
<tr>
<th>D</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer[Cj]</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>
References


- www.aegonreligare.com
- http://www.bankofindia.co.in/english/swasthya.aspx
- http://www.bharti-axagi.co.in/
- www.cholainsurance.com
- http://www.cignattkinsurance.in
- https://epolicy.sblife.co.in
- www.hdfcergo.com
- http://www.medindia.net/patients/insurance/health-insurance-companies/health-insurance-companies-united-states.htm
- http://www.royalsundaram.in/Health-Insurance.aspx
- www.starthealth.in
- www.theorientalinsurance.com
- http://www.webcrawler.com/info.wbcrawl.302.03/search/web?q=health+insurance+providers+list

Introducion

Bibliography

It can be seen from the table that the most recent three years have witnessed a surge in the retail sector in India. The industry is dominated by large retailers such as Reliance Retail, Future Group, and The Tata Group. The industry is witnessing a trend towards consolidation, with many small retailers merging to form larger entities. This trend is expected to continue in the future as retailers look to increase their market share and profitability. The industry is also seeing an increase in the use of technology, with many retailers adopting new technologies to improve their customer experience and increase sales. The retail sector is expected to continue to grow in the future, driven by the increasing purchasing power of the Indian consumer and the expansion of the middle class.