Journal of Autonomous Intelligence (2024) Volume 7 Issue 5
doi: 10.32629/jai.v7i5.1469

ORIGINAL RESEARCH ARTICLE

CSRM-framework for generating actionable knowledge for Social Security Schemes with a special focus on Ayushman Bharat

P. Sunil Kumar*, R. Raghunatha Sarma

Sri Sathya Sai Institute of Higher Learning, deemed to be University, Puttaparthi, Andhra Pradesh 515134, India

*** Corresponding author:** P. Sunil Kumar, psunilkumar@sssihl.edu.in

ABSTRACT

Data Mining applications to Social Security Schemes (SSS) have been one of the most interesting research areas in the recent past. In general, the benefits of SSS are availed by the beneficiaries at different geographical locations at different points in time, thus generating sequential patterns of interest for the stakeholders such as Government bodies or financial institutions to make effective decisions. Typically, SSS launched in the domains of the health sector involves temporal data, and the research in this domain is termed social security data mining (SSDM), where techniques such as sequential pattern mining, sequential rule mining, and association rule mining are in vogue. In this regard, we have proposed a novel data mining framework called the Combined Sequential Rule Mining framework (CSRM-Framework) which is effective in bringing out the actionable knowledge through the activity sequences pertaining to the beneficiaries. The proposed framework was implemented on Ayushman Bharat-Pradahn Manthri Jan Arogya Yojana (AB-PMJAY), a flagship social security scheme launched by the Government of India. We have also proposed a new interesting measure namely Combined Cumulative Lift (CCL) which has the property of estimating the 'Interestingness' effectively when activity sequences are combined with characteristic beneficiary data in the context of AB-PMJAY.

Keywords: combined sequential rule mining; social security schemes; Ayushman Bharat-Pradhan Mantri Jan Arogya Yojana (AB-PMJAY)

ARTICLE INFO

Received: 5 December 2023
Accepted: 26 December 2023
Available online: 30 May 2024

1. Introduction

Lonbing Cao[1] suggested several data mining goals and challenges in the domain of social security to analyze the beneficiary-stakeholder behavior towards the various services provided by Government agencies. Some of the broad categories of analysis that can be carried out in SSDM domain as stated in Cao[1] include customer-centric, process-centric, fraud-centric, and policy-centric analysis. Cao[1] also suggested various challenges like pre-processing, pattern analysis, and knowledge discovery for heterogeneous agents involved in such systems.

"Ayushman Bharat-Pradhan Mantri Jan Arogya Yojana" (PM-JAY), launched in India in September 2018, is one of the largest social security schemes, in the world in the health sector. The scheme intends to move the nation closer to 'Universal Health Coverage (UHC) to achieve Sustainable Development Goals (SDG) by providing health coverage of Rupees 5 lakh per family per year with a minimal premium, for secondary and tertiary care hospitalization. It plans to cover 10.74 crore poor and deprived rural families with special concern for girl children and senior citizens (covering approximately 50 crore beneficiaries). The scheme includes 1354 medical packages covering surgery, medical and day-care treatments of medicines, and diagnostics along with state-wide pre-existing diseases. The scheme is to be jointly implemented by central and state governments with a funding pattern of 60:40 respectively, through national and state health agencies (NHA, SHA) in any of the three modes e.g., Insurance mode/Trust mode/Mixed mode adhering to "standardized treatment guidelines" (STGs) and standardized package rates which promises to build a holistic healthcare ecosystem. Being an entitlement-based scheme, the beneficiaries are identified from rural as well as urban areas and services are delivered based on deprivation criteria. For rural areas, the deprivation criteria include families belonging to five different categories (D1 TO D5), and for urban areas, families belonging to 11 occupational categories are included in the scheme as specified in Liu[2]. It has a number of activities carried out in a sequential fashion giving rise to a large amount of temporal data at various levels of the scheme and various points of service, throughout the country. Typically, an enrolled beneficiary of this scheme gets treatment in an empaneled hospital by undergoing different tests conducted by the parent hospital or referral hospital at different points in time giving rise to the temporal data. The current paper focuses on the problem of predicting the next probable procedure or package availed by a beneficiary and also provides actionable knowledge for the stakeholders by using CSR-Framework.

The uniqueness of our technique lies in combining the transactional data of AB-PMJAY with the characteristic data of beneficiaries to generate actionable rules. The transactional data results in activity sequences of beneficiaries due to multiple services availed under the scheme. The positive impact of the proposed interesting measure namely Combined Cumulative Lift (CCL) is that it has the potential to reduce the overall cost borne by the stakeholders and also the waiting time of the beneficiary in availing the services under PM-JAY.

2. Materials and methods

2.1. Sequential pattern mining

Sequential pattern mining has been one of the major domains of research in the recent past. The research activities comprise mainly of sequential prediction, sequential rule mining and pattern mining. Normally these mining techniques generate huge number of patterns for large transaction data sets. In order to overcome this, Zhang and Wu deliberated about the need of frequent patterns that are easy to analyze[3,4]. Most of the sequential pattern mining algorithms like SPAM, SPADE, and GSP were based on positive sequential rules only, as suggested by Fournier-Viger et al.[5]. Zhu et al.[6] suggested an algorithm that uses tree projection to find all the frequent patterns for a predefined window length and threshold. The input sequences are given for finding all the frequent patterns using the algorithm. However, for domain-specific applications, the algorithm suggested in Zhu et al.[6] requires prior knowledge of all possible procedures and their corresponding mappings to the frequent patterns, which must be known in advance. Rezig et al.[7] worked on predicting the sequence of maintenance activities within a time window of 3–4 weeks. Rezig et al.[7] has used the GSP algorithm to identify the frequent sequential patterns of all lengths followed by an a-priory algorithm to generate the sequential association rules for all the maintenance activities. Two important measures viz. coverage and accuracy were suggested by the author to predict the next maintenance activity along with the frequent spare parts involved in the sequence of activities. However, Rezig et al.[7] does not focus on the order and length of the sequences during that time window. Fournier-Viger et al.[8] has worked on predicting the next item set in a sequence of symbolic sequences viz; browsing web documents. The authors used partial ordered sequential rules instead of the standard sequential rules to improve the accuracy under several scenarios. The notions such as prefix-size and suffix-size are used to represent the lengths of antecedent and consequent respectively in the rules thus overcoming the limitations of order and length of the sequences. However, the number of rules that are generated after matching the sequence for a new input is not pruned efficiently. Zhao et al.[9,10] suggested negative sequential rules along with positive sequential rules for sequential classification tasks. However, the work carried out by them has limitations about timing-window as well as generated rules. Wright et al.[11] worked on the sequential patterns of diabetic patients obtained from their prescriptions at the generic drug level as well as a drug class. The authors used the cSPADE algorithm and generated rules to predict the next temporal relationship between the medications based on brief patient history. The work assumed that the last medication in the sequential database is treated as the consequent and the items preceding it as the antecedent for all the rules thus limiting the length for antecedents and thus restricting the intermediate intervention during the medication. Zhao et al.[12,13] and Zhu et al.[6] suggested combined association rule mining to extract actionable knowledge in the social security domain. These rules are added with some interesting measures to provide actionable knowledge by combining the demographic data with the transaction data. This process reduced the number of association rules to a large extent.

The proposed work is to provide actionable knowledge to the stakeholders by using combined sequential rules as the data involves temporal data. In this approach, the rules are generated in two steps. Firstly, sequential rules are generated from a universal sequence by mapping onto the transaction database. Secondly, the beneficiaries are grouped based on certain criterion like age, gender, etc. and finally, they are combined to generate rules that ideally involves techniques like pruning and reduction of search space suitable for AB-PMJAY.

2.2. Combined sequential rule framework

The proposed Combined Sequential rule framework (CSR-Framework) is shown in **Figure 1**. For any typical social security scheme launched by the Government, the scheme-related information and its benefits were disseminated to the eligible beneficiaries by using broadcast model using different Information and Communication Technologies (ICT) as described by Khalid[14]. The information related to services provided at different government units at different geographical locations, beneficiary details, and their feedback is collected using service delivery models and lobbying models.

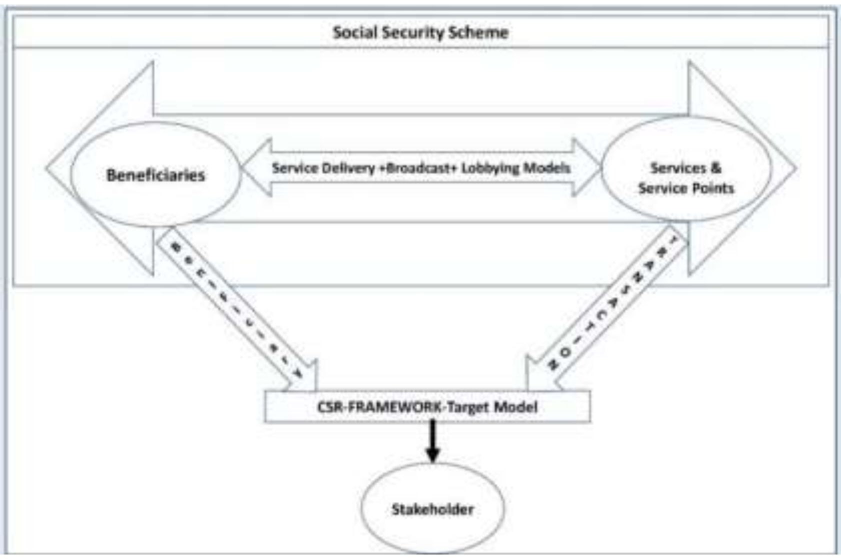

Figure 1. Social security scheme-CSR framework.

The stakeholders for such schemes have to make crucial decisions at various points of action depending upon the feedback from various class of beneficiaries as well as their transaction data. Instead of taking the feedback separately, they are combined into the proposed CSR-framework which provides a set of actionable rules for each class of beneficiaries which makes the decision process easy for the stakeholders. This is represented as the target model or crucial workflow model as suggested by Khalid[14].

Figure 2 shows the combined sequential rule generation framework for AB-PMJAY scheme. The designated beneficiaries of the rural as well as urban areas were identified and awareness about the scheme benefits was provided using the standard e-governance models. The transaction database of the registered beneficiaries was explored to generate sequential rules.

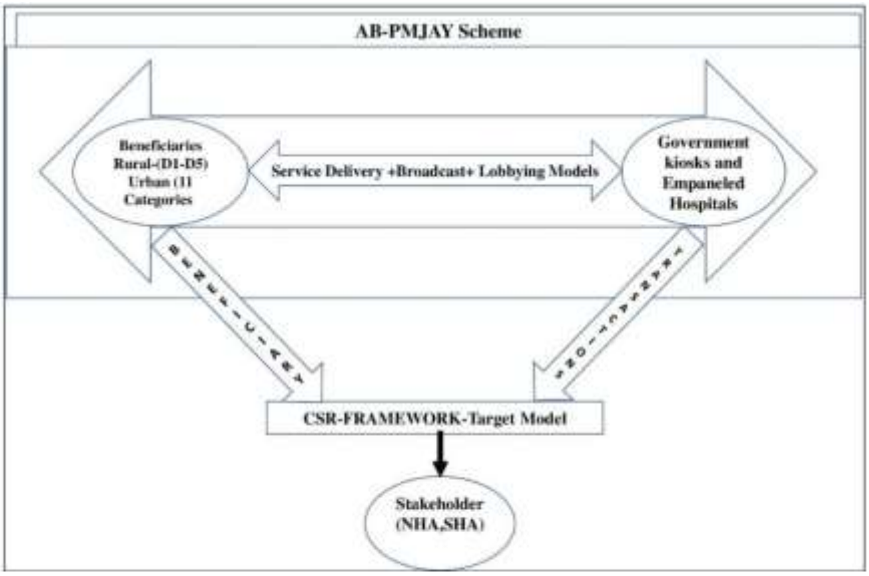

Figure 2. ABPMJAY-CSR framework.

Then for different aggregated classes of the beneficiaries belonging to a particular region and deprivation criteria, combined sequential rules were generated using the CSR-Framework.

2.3. Mathematical preliminaries for combined sequential rules

The social security framework can be visualized as a sequence of activities giving rise to a sequence database. A sequence database in this domain is realized from the transaction database of the beneficiary. The transaction database contains temporal information about the beneficiary who visits various points of service to avail the benefits of the scheme.

Let $S = \{s_1, s_2, s_3, \dots s_n\}$ be a sequence database and $A = \{A_1, A_2, A_3, \dots A_m\}$be a set of activities where each sequence s_i is of the form $s_i = \{P_1, P_2, P_3, \dots P_k\}$ such that $P_1, P_2, P_3, \dots P_k \subseteq A$.

Let $C = \{C_1, C_2, C_3, \dots C_l\}$ be a set of classes where each beneficiary belongs to a particular class $C_l \epsilon C$. C_l generally, representing aggregated data that identifies a group of beneficiaries based on a certain criterion of interest. For example, the criteria could be demographic information such as gender, age, place of in-habitation, income range etc.

A sequential rule which is represented as Antecedent → Consequent is given by $\{B\} \to A_j$ where the set B denotes any ordered sequence of A_j's where $A_j \in$ A. In the context of social security schemes, each sequential rule of the form $\{B\} \to A_j$can be interpreted as an immediate occurrence of an activity A_j after the occurrence of sequence {B}. A combined sequential rule is defined as $(C_l \wedge \{B\} \to A_j)$, where the antecedent of the combined rule represents all the sequences that satisfy the class criteria.

Hence a combined sequential rule can be interpreted as an immediate occurrence of an activity A_j after the occurrence of the sequence{B} for all those beneficiaries belonging to the class C_l.

The interesting measures for the combined sequential rules are listed in **Figure 3**.

Interesting Measures for the Combined Sequential rules of the form $C_l \wedge \{B\} \rightarrow A_j$
Support: $$\frac{Support(C_l \cup \{B\} \cup A_j)}{\lvert S \rvert}$$
Confidence: $$\frac{Support(C_l \cup \{B\} \cup A_j)}{Support(C_l \cup \{B\})}$$
Lift: $$\frac{Support(C_l \cup \{B\} \cup A_j)}{Support(C_l \cup \{B\}) \times Support(A_j)}$$

Figure 3. Interesting measures.

2.4. Application of CSRM framework in AB-PMJAY

The role of CSR-Framework in the AB-PMJAY scheme relates to an eligible beneficiary residing in rural as well as urban areas satisfying the conditions laid down by the scheme. An enrolled beneficiary visits an empaneled hospital to avail the benefits. Depending upon the illness the beneficiary is redirected to one of the specializations where he or she undergoes a set of procedures at different points of time which constitutes the transaction data for the proposed framework.

As an initial step, the universal set of procedures availed by all the beneficiaries for a specific specialization at an empaneled hospital is identified. Then the domain experts of the hospital are consulted to understand the frequently availed sequences of procedures by the beneficiaries. It was concluded after the discussion with the experts that the treatment sequences are of finite length. So, the current work focuses on treatment sequences of length three only. Sequential rules are first generated for these sequences to obtain the frequently used procedures. The beneficiary class C_l is formed by aggregating the demographic details and the procedures from the activity sequences under the AB-PMJAY scheme.

In traditional sequential rule mining, one encounters the problem of rule explosion, i.e., a large number of sequential rules are generated which can be unwieldy. With combined sequential rule mining, the number of rules generated is curtailed due to the class constraint imposed on the beneficiaries.

2.5. Data

AB-PMJAY scheme provides benefits to the enrolled beneficiaries in 23 different specializations as ruled out by National Health Authority of India (NHA) specified in HBP[15]. The empaneled hospitals offering treatment in any of the specializations under the scheme are required to meet the criteria set by the stakeholders about infrastructural and treatment facilities. Each specialization in turn comprises a set of packages or procedures used for diagnosing and treating the affected beneficiaries. In addition to these specializations, the state-specific specializations ruled out by respective State Health Authorities (SHA) are also included for the beneficiary's benefit. For our research work, the data was collected through a survey done on registered beneficiaries of AB-PMJAY visiting Sri Sathya Sai Higher Medical Sciences (SSSIHMS), Prasanthinilayam for treatment. The data set includes the treatment history of 534 patients who made multiple visits at different points in time under the Investigation department. The survey outcomes also include various demographic inputs about the beneficiaries such as their age groups, derivation criteria, and their domicile information such as rural or urban, etc. It was also found that the beneficiaries have undergone 9 different procedures in common but in different order. These procedures are CGHS-ECG, CGHS-Chest PA view (one film)" CGHS-2D echocardiography, CGHS-Biopsy, CGHS-Occlusal X-ray, CGHS-Stress Echo-exercise, CGHS-Abdomen

Lateral view (one film), CGHS-Chest Lateral (one film) and CGHS-Upper G.I. Endoscopy Lower G.I. Endoscopy. These procedures constitute the universal set A. For analysis purposes, the procedures are encoded as P1–P9. A typical snapshot of the data collected is presented in the following **Table 1**. The data shown here is hidden for security purposes. It is important to note here that the redundant information was removed by adopting data quality checks and aggregation techniques. The missing information from the transaction dataset are handled using KNN.

Table 1. A snapshot of the transaction data of the beneficiaries.

Arrival	Patient Id	Age	Gender	Procedure	Place
2020-12-01 11:50:00	XXXX 549	67	Female	P1	Rural
2020-12-01 11:50:00	XXXX 549	67	Female	P3	Rural
2020-12-01 11:50:00	XXXX 549	67	Female	P6	Rural
2021-06-30 11:20:00	XXXX 539	73	Female	P1	Rural
2021-06-30 11:20:00	XXXX 539	73	Female	P4	Urban
2021-06-30 11:20:00	XXXX 539	73	Female	P3	Rural
2021-06-30 11:20:00	XXXX 539	73	Female	P2	Rural
2021-02-27 17:21:00	XXXX 817	69	Male	P1	Urban
2021-02-27 17:21:00	XXXX 817	69	Male	P3	Urban
2021-02-27 17:21:00	XXXX 817	69	Male	P2	Urban
2020-09-30 18:13:00	XXXX 207	55	Female	P1	Urban
2020-09-30 18:13:00	XXXX 207	55	Female	P3	Rural
2020-09-30 18:13:00	XXXX 207	55	Female	P2	Rural

2.6. Implementation of CSRM framework

Table 1 shows a snapshot of the patients along with their demographic details. To provide actionable knowledge to the stakeholders using the CSR Framework, the major steps being followed are:

- Step-1: Generating the sequential rules for the transaction data.
- Step-2: Generating the combined rules by combining the aggregated data and transaction data.
- Step-3: Interesting measures applied to combined sequential rules

2.6.1. Generating the sequential rules for the transaction data

The sequence data is generated from the patient visiting data, as shown in **Table 2.**

Generate and test: For the given set of procedures, all possible sequences of length one to three or more are generated and mapped onto the sequence data generated in the previous step to get the count of each sub-sequence required for the interesting measures.

Table 2. A snapshot of Sequence data.

Patient Id	Age	Gender	Procedure Sequence	Place
XXXX549	67	Female	’P1’,’P3’,’P2’	Rural
XXXX539	73	Female	’P1’,’P4’,’P3’,’P2’	Rural
XXXX817	69	Male	’P1’,’P3’,’P2’	Urban
XXXX207	55	Female	’P1’,’P3’,’P2’	Rural

In the current problem, each patient undergoes a treatment sequence independent of other patients’ sequence concerning limit or order. In our problem, the sequence {1,2,3} is different from {2,3,1}.

For each sequence obtained in the previous step the sequential rules of the type $\{B\} \rightarrow A_j$ were generated along with their interesting measures, where Antecedent {B} is the sequence representing the current treatment of the new patient. The consequent A_j predicts the next possible procedure in the given sequences e.g., for j = 1, 2, 3, 4, 5, 6, 7, 8, 9.

Support is the probabilistic measure of the number of transactions where {B} and A_j occur together. Confidence is the probabilistic measure of the number of transactions where A_j occurs after every occurrence of {B}. Lift is a measure of the interestingness of the rule. Lift of $\{B\} \rightarrow A_j$ is a probabilistic measure of the confidence of the rule to the expected confidence of the rule. If lift is greater than or equal to 1 then the rule is said to be interesting. In this paper, more focus is given to the lift measure which is obtained from the support and confidence measures of each sequential rule.

Table 3 shows lift measures for a few of the most frequently occurring rules of the form $\{B\} \rightarrow A_j$, obtained from the transition data.

Table 3. A snapshot of sequential rules with interesting measures.

$\{\boldsymbol{B}\} \rightarrow \boldsymbol{A_j}$	$\boldsymbol{Lift}\ \{\boldsymbol{B}\} \rightarrow \boldsymbol{A_j}$
P1$\rightarrow P$ 3	0.97
P1$\rightarrow P$ 4	0.94
P1$\rightarrow P$ 2	0.92
P1, P3$\rightarrow P$ 2	0.94

2.6.2. Generating the combined rules by combining the aggregated data and transaction data

The demographic data is aggregated based on the age, place, and gender of the patients into 16 classes as shown in **Table 4**.

Table 4. Aggregated classes of Beneficiaries.

Group	Gender	Age Interval	Place
C1	Female	1–18	Rural
C2	Female	19–35	Rural
C3	Female	35–60	Rural
C4	Female	More than 60	Rural
C5	Female	1–18	Urban
C6	Female	19–35	Urban
C7	Female	35–60	Urban
C8	Female	More than 60	Urban
C9	Male	1–18	Rural
C10	Male	19–35	Rural
C11	Male	35–60	Rural
C12	Male	More than 60	Rural
C13	Male	1–18	Urban
C14	Male	19–35	Urban
C15	Male	35–60	Urban
C16	Male	More than 60	Urban

The class of beneficiaries is denoted as C_l. A snapshot of the aggregated groups along with the sequence data shown in **Table 5**. For each class of beneficiary C_l sequential rules of the form $\{C_l \rightarrow A_j\}$ is generated along with their interesting measures like support, confidence, and lift shown in **Table 6**. This step helps in analyzing the treatment sequence of different age groups of patients. In this case A_j can be any one of the 9 procedures under study.

Table 5. A snapshot of transaction data along with aggregated classes.

Patient Id	Category	Procedure Sequence
XXXX549	C4	'P1','P3','P2'
XXXX539	C4	'P1','P4','P3','P2'
XXXX817	C16	'P1','P3','P2'
XXXX207	C7	'P1','P3','P2'

Table 6. A snapshot of Interesting Measures for three different classes.

Class	Patients	Patients with P2	Confidence	Lift
C3	94	82	0.87	0.83
C4	34	27	0.76	1
C8	34	26	0.79	0.82

For each sequence obtained in the previous step, the combined sequential rules of the type $(C_l \wedge \{B\} \rightarrow A_j)$, are generated along with their interesting measures by combining the aggregated demographic data with the current treatment sequence.

Here the antecedent is a pair $(C_l \wedge \{B\})$, denoting the current treatment sequence by a particular class of beneficiary and the consequent is the probable procedure to be provided for that particular class is considered. **Table 7** shows a summary of interesting measures obtained after calculating the interesting measures based on transaction data and the aggregated information in the form of classes for three major classes of the survey namely C3, C4 and C8. The aggregated information taken for different classes of the beneficiaries based on their demographic data along with the frequent sequential rules for the given transaction data, are the two major dimensions based on which combined sequential rules are framed. The aggregated information for this work was taken after taking the expert opinion into consideration.

Table 7. A snapshot of Combined Lift for the Three classes C3, C4 and C8.

$(\{B\} \rightarrow A_j)$	(C3 ∧ {B})	(C4 ∧ {B})	(C8 ∧ {B})
P1$\rightarrow P\,3$	1.08	1.07	1.19
P1$\rightarrow P\,4$	0.96	1.00	1.08
P1$\rightarrow P\,2$	1.03	1.04	1.04
P1, P3$\rightarrow P\,2$	1.03	1.17	1.73

2.6.3. Interesting measures applied to combined rules

{Combined Cumulative-lift (Class-wise)}: This measure calculates the contribution of demographic information over the current sequence of treatment to generate actionable knowledge in the combined rule.

In this formula {B} represents the set of all possible procedures as sub-sequences availed by the patient and A_j represents the next procedure that the patient expected to undergo.

Let Total_Sequnce_based _Lift = $\sum_{n=1}^{k}(\text{Lift }(\{\text{B}\}) \rightarrow A_j))$

where 'k' represents the number of classes under consideration and 'n' represents the number of all possible classes.

Combined_Cumulative_Lift (Class-Wise) is given by

$$\prod_{n=1}^{k} \frac{Lift((C_l \wedge \{B\}) \rightarrow Aj)}{\text{Total\_Sequnce\_based \_Lift}}$$

{Combined Cumulative-lift (Sequence-wise)}: This measure calculates the contribution of current sequence of treatment for generating actionable knowledge in the combined rule.

Let Total_Class_based_Lift = $\sum_{n=1}^{k}(\text{Lift}(\{C_l \rightarrow A_j\}))$

Combined_Cumulative_Lift (Sequence-Wise) is given by

$$\prod_{n=1}^{k} \frac{Lift((C_l \wedge \{B\}) \rightarrow Aj)}{\text{Total\_Class\_based \_Lift}}$$

3. Results

Out of 3609 possible sequences for 9 procedures, 511 sequences with sequence length 3 are chosen for the analysis purpose.

The patients belonging to the groups C3, C4, and C8 are being analyzed in this study whose occurrence in the data set is shown in **Table 6**. Usage of the procedure patterns by this class of beneficiaries is considered. The last procedure being availed by this class of beneficiaries was found to be {P2} in most of the patterns. Hence the sequential rules of the form C_l → P2 as well as their interesting measures were calculated which are shown in the **Table 5**.

The combined sequential rules are shown in **Table 7**. It can be observed that the lift measures of the sequential rules raised considerably after combining the demographic information. It can be seen that for the sequential rule ($\{B\} \rightarrow A_j$), where B= {P1, P3}. and A_j={P2} the initial lift measure was 0.9482. Its value increased to 1.0,1.17 and 1.73 after combining the group information e.g., C3, C4, and C8 respectively. It can be understood that females aged 60 or above living in rural or urban areas are expected to undergo the CGHS-Chest PA view (one film)" CGHS-2 D echo cardiograph if they have undergone the procedures CGHS-ECG, CGHS-Biopsy in order. Procedure usage frequency was very much evident from the results for a given class of beneficiaries.

After applying interesting measures on the combined sequential rules as shown in **Table 7** for the three classes C3, C4 and C8 with A_j = P2 and the sequential rule ($\{B\} \rightarrow A_j$), where B = {P1, P3} and A_j = {P2}. We obtain Cumulative-Lift (Class-Wise) = 2.19 and Cumulative-Lift (Sequence-Wise) = 1.53 which shows that the combined sequential rules with the demographic information provide more insight and gain in actionable knowledge by the stakeholders for effective decision-making.

The health benefit packages under AB-PMJAY are revised based on the usages at the impaneled hospitals from time to time by the stakeholders as published at HBP[15]. The latest modifications were done in November-2021. The suggested framework helps the stakeholders to identify the frequently used procedures for a given class of beneficiaries. For the survey conducted in this study, it was found that the procedures CGHS-ECG, CGHS-Chest PA view (one film)" CGHS-2D echocardiography, CGHS-Biopsy were more frequently used than others. The framework also helps the doctors and stakeholders to have early information about the most affected class of beneficiaries and treatment sequences at any specific demographic zone using CSR-Framework used in this paper.

It was very much evident from the obtained results is that the female beneficiaries within the age group 35–60 belonging to rural as well as urban areas, have to undergo CGHS-Chest PA view (one film)" CGHS-2D echo cardiograph if they have undergone CGHS-ECG or CGHS-Biopsy.

In the proposed framework the interestingness measures shown in **Table 7** for the combined sequential rules improved the accuracy in comparison to **Table 6** while predicting the next procedure for a given class of beneficiary. In addition to this, as the AB-PMJAY scheme was devised to reduce the out-of-pocket expenses borne by the beneficiaries, the framework also supports this regard by providing the next most probable procedure to be undergone by the beneficiary and hence reduce the overall cost.

4. Conclusion

It is appropriate to mention that the CSRM-framework is useful under any Government initiated social security schemes that involves beneficiary, benefits and transactions at different locations at different points of time giving rise to temporal data. Since these schemes generate large scale of data on a daily basis the division of analysis comprising of people and transactions using CSRM-framework will be apt for the stakeholders from business point of view in order to make long term decisions for differed class of beneficiaries namely senior citizens, children and youth etc.

For a given set of procedures or activities, the dataset chosen for the said work requires transaction data, with multiple occurrences of the procedure or items but without any repetition. Secondly the current study assumes that the length of the procedure-sequence should not exceed 3, which was suggested by the domain experts. Handling the sequences with larger length will be a complex task. Further research activity can be carried out to address such issues.

Author contributions

Conceptualization, PSK and RRS; methodology, PSK; software, PSK; validation, PSK and RRS; formal analysis, PSK; investigation, PSK; resources, RRS; data curation, PSK; writing—original draft preparation, PSK; writing—review and editing, RRS; visualization, PSK; supervision, RRS; project administration, RRS; funding acquisition, RRS. All authors have read and agreed to the published version of the manuscript.

Acknowledgments

We express our deep sense of gratitude to the founder chancellor of our university, Bhagawan Sri Sathya Sai Baba for His inspiration and message for life which becomes the ultimate purpose of this proposed work.

Conflict of interest

The authors declare no conflict of interest.

References

1. Cao L. Social security and social welfare data mining: An overview. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2012, 42(6): 837-853. doi: 10.1109/tsmcc.2011.2177258
2. NHA. National Health Authority. Available online: https://nha.gov.in/PM-JAY.html (accessed on 21 November 2022).
3. Wu S, Zhao Y, Zhang H, et al. Debt Detection in Social Security by Adaptive Sequence Classification. Lecture Notes in Computer Science. Published online 2009: 192-203. doi: 10.1007/978-3-642-10488-6_21
4. Zhang H, Zhao Y, Cao L, et al. Customer Activity Sequence Classification for Debt Prevention in Social Security. Journal of Computer Science and Technology. 2009, 24(6): 1000-1009. doi: 10.1007/s11390-009-9288-2
5. Fournier-Viger P, Lin CW, Rage U, et al. A survey of sequential pattern mining. Data Science and Pattern Recognition. 2017, 1: 54–77.
6. Zhu C, Zhang X, Sun J, Huang B. Algorithm for mining sequential pattern in time series data. 2009.
7. Rezig S, Achour Z, Rezg N. Using Data Mining Methods for Predicting Sequential Maintenance Activities. Applied Sciences. 2018, 8(11): 2184. doi: 10.3390/app8112184
8. Fournier-Viger P, Gueniche T, Tseng VS. Using Partially-Ordered Sequential Rules to Generate More Accurate Sequence Prediction. Lecture Notes in Computer Science. Published online 2012: 431-442. doi: 10.1007/978-3-642-35527-1_36

9. Zhao Y, Zhang H, Wu S, et al. Debt Detection in Social Security by Sequence Classification Using Both Positive and Negative Patterns. Lecture Notes in Computer Science. Published online 2009: 648-663. doi: 10.1007/978-3-642-04174-7_42
10. Zhao Y, Zhang H, Cao L, et al. Efficient mining of event-oriented negative sequential rules. 2008.
11. Wright AP, Wright AT, McCoy AB, et al. The use of sequential pattern mining to predict next prescribed medications. Journal of Biomedical Informatics. 2015, 53: 73-80. doi: 10.1016/j.jbi.2014.09.003
12. Zhao Y, Zhang H, Cao L, et al. Combined association rule mining. 2008.
13. Zhao Y, Zhang H, Figueiredo F, et al. Mining for combined association rules on multiple datasets. 2007.
14. Fakieh K. The E-Governance (E-GOV) Information Management Models. International Journal of Applied Information Systems. 2016, 11(1): 10-14. doi: 10.5120/ijais2016451567
15. HBP. Health Benefit Pacakages. Available online: https://nha.gov.in/img/resources/ HBP-2.2-manual.pdf (accessed on 1 November 2021).