Health Insurance Data Processing using Linked Data

Vishal Shah, Shridevi S

Abstract: ***The data in the organization is distributed among multiple structured databases. The large database makes the process of risk analysis difficult, as data is distributed in the organization. Information gathering for Risk analysis is more subjective and therefore, processing incomplete information over distributed databases increase more fault in risk analysis. Linked Data representation helps to make structured, distributed data more related, combined and ready to be processed. Linked Data approach makes data interlinked and semantically rich, extracting meaning with the use of machines and eliminating the human subjectivity factor in assessing insurance risk. Using Linked data, information retrieval process can be easier as data or databases interlinked semantically. The proposed technique uses a linked data approach for risk analysis and related information retrieval methods over structured data. The work efficiency is also tested and found to be good.***

Keywords: Linked Data Analysis, Semantic-web, Ontology, Information retrieval and Gathering.

I. INTRODUCTION

Advancement in data gathering and in data enrichment at an industrial level increase day by day. And it is necessary to improve the quality of data for the diverse business process involved. The blast of data is driven by two specific sources: the interpersonal network sharing information about our activities and an assortment of data recovery form collide information on our condition. Data gathering is a big challenge that minimizes data scarcity in an organization. Structured data storage in the organization has its own limitations. Advanced analytics require external data that gives additional information to make a prediction in a better way. BigData problems in insurance industries are still there compared to other industries. As disparate datasets define limits in insurance industries, Linked data representation helps in overcoming that. Life insurance industry more often has a complex structure of data where user personal details are stored in relational databases in distributed places. Through advanced analytics, we can find risk factors sources in the life insurance sector. Such possible risk factors are [3]:

1. Overall health and pre-existing medical conditions
2. High-risk activities
3. Occupation
4. Lifestyle factors
5. Lifestyle habits such as smoking or drinking alcohol

Linked data approach gives the opportunity to make data more connected with external data so that advanced analytics will be more accurate. Linked data technologies define great opportunities to make data interlinked with external data resources [1]. They use the resource description framework (RDF) language and HTTP protocol to publish structured data on the Web [2]. Linked data ontologies can be shared nowadays openly so that more useful methods applied to ontologies. Linked data over structured data in an organization is a challenging task but under particular ontologies can be developed. Therefore, connected ontologies can be analyzed in a more proper fashion and advanced analytics problems can be solved.

Information retrieval and gathering make linked data produce more useful result thereby making the decision easier. This paper explores the approach for converting structured data to linked data for analyzing risk by combining external data ontologies to the existing one to identify.

Further, in this paper, we have following sections. Section 2- Related work gives previous research related to linked data. Research questions give a review of our literature. Then, a proposed methodology that gives information about methods and techniques we are using. Finally we conclude with results-discussion and conclusion section.

II. RELATED WORK

Various research works have been completed to enhance the quality of data using a linked data approach. Many studies give an idea of combined approaches or real-world applications where structured data problem are solved using linked data representation. The below subsections discusses few domains where linked data approach helped in reaching better results.

A. Linked Data, Datamining and External Open Data for Better Prediction of At-Risk Students

Survey-based study more often time-consuming and the data correctness problem is there while we process that data. Students interest in their higher studies is important nowadays and it is the biggest challenge to improve student retention [4,5]. Data mining techniques are applied to study the patterns from past data of student database and models have been developed for students retention during recent years. One of the commonly used model is Tinto's model [6,7], where the likelihood of a student withdrawing from higher education is seen as being determined by individual attributes, familial attributes, prior qualifications, social integration, academic integration, individual commitment, institutional commitment, and external family and societal factors.

Manuscript published on 30 August 2019.
*****Correspondence Author(s)

Vishal Shah, SCSE, Vellore Institute of Technology, Chennai, India.
Dr. Shridevi S, SCSE, Vellore Institute of Technology, Chennai, India.

The dataset which contains multiple attributes has to be studied well to make a model which identifies students at-risk. Various sources of databases are institutional internal databases and external open data. But, the challenge is to get that database integrated [8].

Therefore, the linked data approach was utilized and with the neural network identified map attributes with connected data that achieved an overall model accuracy of 84.94% [9].

B. Framework for The Identification of Fraudulent Health Insurance Claims using Association Rule Mining

Fraudulent claims of insurance incur a heavy loss to the insurance companies and from recent studies, it is known that 15% of the claims of insurance are fraudulent [10]. Fraudulent claims lead to different ways of claiming insurance money and thus it increases forged claims to the insurance company. The way to identify fraudulent claims can be done thru data mining techniques. Supervised and unsupervised data mining technique has their own benefit in processing the data to produce better results. The supervised technique works with train and test data. And the unsupervised technique able to discover patterns from existing data and new data also. Here, authors have completed a study of a mixture of both the techniques. With unsupervised technique, identify patterns and then with the supervised technique, based on that patterns, different classes are identified from the dataset.

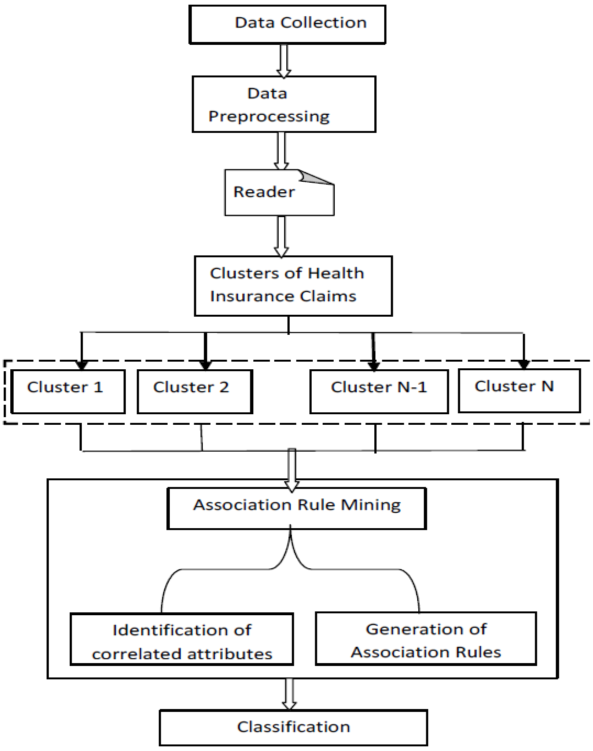


Fig 1: Framework for Identify Fraudulent Insurance Claims

By this approach, authors also presented association rule mining as a promising and best fitting approach to finding fraudulent claims within the health insurance domain [11].

C. How Structured Data (Linked Data) Help in Big Data Analysis – Expand Patent Data with Linked Data Cloud

Big Data is the solution for handling or managing the large structured or unstructured database processing. Big data is described when data in conventional databases exceeds the processing capacity in systems [12]. Big Data can be structured or unstructured with the four main characteristics Volume, Variety, Velocity, and Volume. The main challenge with Big Data is that processing of that data with efficient management and algorithmic techniques [13]. Where on the other hand Linked data approach allows integrating interrelated datasets that can be published and shared on the web. Day by day the linked data open cloud is growing constantly, that means data integration is becoming more important in this field.

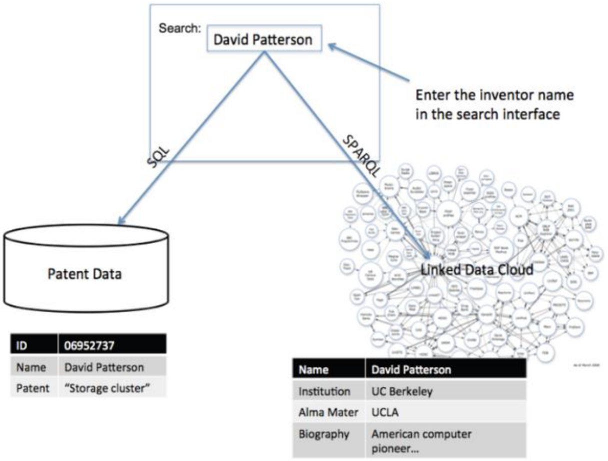


Figure 2: The Querying Process of Patent Search Engine

In this project [14] authors have built a Patent Search Engine where large dataset of patent holders and their information stored on SQL database. From that SQL database, the particular information for the patent holder is collected and correlated with open linked data cloud. Using that linked data, SPARQL (SPARQL Protocol and RDF Query Language) query is applied to collect all interlinked data for the user query. Thus linked data approach benefits information retrieval for user queries.

D. A Linked Data Approach for Geospatial Data Provenance

Geospatial data provenances are used to obtain geoinformation from the cloud easily. Linked data cloud grows day by day constantly by linking useful information with that cloud. Linked information provides semantically rich content and also it's machine-readable as linked data allow interlinking data based on its characteristics. This paper proposes a catalog for publishing data into linked data cloud [15].

III. REASEARCH QUESTIONS

This section gives the purpose of the proposed system by identifying certain research gaps and tries to give solution for that. The below questions are based on this research:

A. What is the Limitation with structured data in the insurance companies?

The limitations with structured data limit insurance companies to do advanced analytics. Many types of research are going on different techniques related to advanced analytics.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

Different factors are there that have to be considered for predicting future analysis.

B. Why linked data approach so beneficial for advanced analytics?

Linked data provides more opportunity for data enhancement, where data is in a linked data model thereby providing data having the characteristic of relativity. Therefore, any futuristic analysis will be easier through linked data.

C. How linked data approach applied for risk analysis in the life insurance industry?

The biggest challenge to dig out in the insurance sector is risk analysis in earlier stages. Insurance industry generating structured data to store all the information of insurer and their claims. RDF modeling of insurer data and application of Rules to that data makes them more interlinked. Data extraction can be done through SPARQL query to get domain centric data.

D. In which context advanced analytics works with linked data in the life insurance industry?

Rule-based mining is applied to the data and then according to the rules, data are segregated based on certain condition High, Low and Medium level risks are identified. Specific Insurer's data can be analyzed with the linked data approach where their other information or geospatial information can also be analyzed.

Now, based on the research questions the proposed system is developed.

IV. PROPOSED METHODOLOGY

The proposed system for this research consists of several important components namely data collection, data preprocessing, association rule mining, categorization of risk levels, D2R server processing, analysis with ontology and classified results. The below diagram illustrates the working of the proposed system.

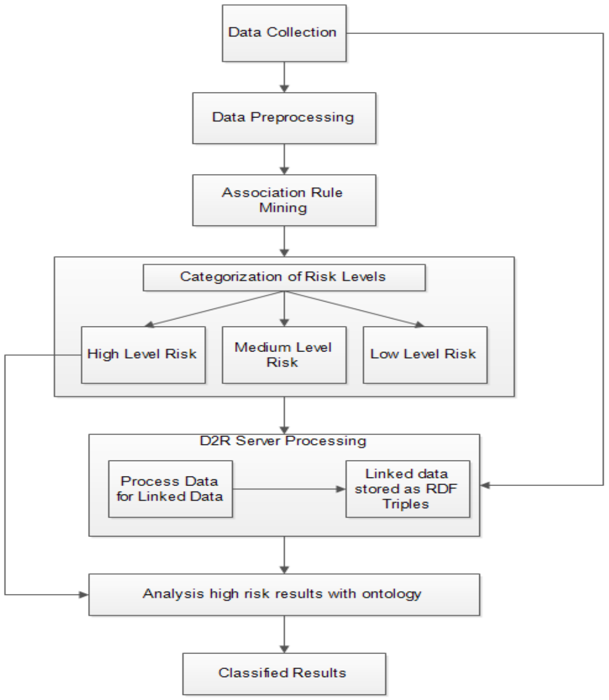


Fig 3: Framework for Advanced Risk Analysis

A. Data Collection

Data collection for this framework completed through various ways like claims data, insurer personal data sets, and geospatial data. Partially the database is collected from kaggle.com [16]. Other related information is combined from different databases. Geospatial data is scratched from online sources through user activities and by tracking their day to day life [17].

B. Data Preprocessing

Data cleaning is the initial step of data pre-handling. It expels irregularities and redundancies from the chose data sets. Data Integrity consolidates the data which is accumulated from various sources at one place e.g. numerous databases, documents. After the data incorporation, the last advance of data pre-handling is performed which is data change. Data change changes over the data into a shape which is required for the examination for instance in this investigation the required data is "health insurance database" so data change will be finished by applying normalization to the data and health insurance claims database will be made.

C. Association Rule Mining

Rule mining in insurance data set is to identifying related attributes that can give frequent rule on the basis of high, medium and low risk level. In the association rule mining attributes identification is the first task that should be initiated first in the algorithmic approach.

- *Attributes Finalization*: The base in association rule mining approach is handling of attributes that relatively find rules for your conditions. Therefore, attributes and only necessary attributes finalization is very crucial in this process.
- *Rules Creation*: Rules are generated using apriori data mining algorithm. Apriori algorithm is used to get frequent itemsets and relevant association rules. Dependency created in association rule e.g. A → B where B is dependent on A for our rule. A and B are two different item sets [18].

D. Categorization of Risk Levels

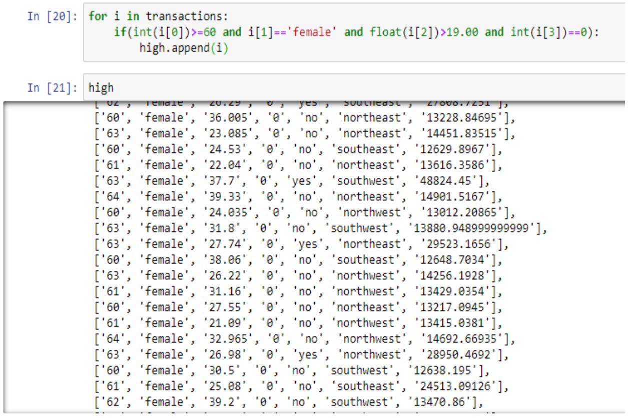
After generating the rules, risk level identified and based on that High, Medium and Low level risk records identified. Those records tagged based on the risk level. The process to identify different risk level depends on the insurance company but here we consider only high-risk measures to identify the high-level risk people.

Let consider certain measures for the high-risk level that insurance company identified based on their past experience.

$$\text{record=>(age}\geq 50 \text{ \&\& BMI}\geq 28 \text{ \&\& region=="southeast" \&\& smoker == "yes" \&\& } charges \;\geq\; 30000) \qquad (1)$$

The above condition for identifying high-risk record is an example to consider. The above equation tells that in particular rules the records matching to that condition are high-risk records. In that equation, that person whose age is greater than 50 and BMI (body mass index) greater than 28 and region are matched to "southeast" and smoker also and having charges greater than 30000 pending.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)
© Copyright: All rights reserved.



```
In [20]: for i in transactions:
             if(int(i[0])>=60 and i[1]=='female' and float(i[2])>19.00 and int(i[3])==0):
                 high.append(i)

In [21]: high
         ['62', 'female', '26.29', '0', 'yes', 'southeast', '27808.7251'],
         ['60', 'female', '36.005', '0', 'no', 'northeast', '13228.84695'],
         ['63', 'female', '23.085', '0', 'no', 'northeast', '14451.83515'],
         ['60', 'female', '24.53', '0', 'no', 'southeast', '12629.8967'],
         ['61', 'female', '22.04', '0', 'no', 'northeast', '13616.3586'],
         ['63', 'female', '37.7', '0', 'yes', 'southwest', '48824.45'],
         ['64', 'female', '39.33', '0', 'no', 'northeast', '14901.5167'],
         ['60', 'female', '24.035', '0', 'no', 'northwest', '13012.20865'],
         ['63', 'female', '31.8', '0', 'no', 'southwest', '13880.948999999999'],
         ['63', 'female', '27.74', '0', 'yes', 'northeast', '29523.1656'],
         ['60', 'female', '38.06', '0', 'no', 'southeast', '12648.7034'],
         ['63', 'female', '26.22', '0', 'no', 'northwest', '14256.1928'],
         ['61', 'female', '31.16', '0', 'no', 'northwest', '13429.0354'],
         ['60', 'female', '27.55', '0', 'no', 'northeast', '13217.0945'],
         ['61', 'female', '21.09', '0', 'no', 'northwest', '13415.0381'],
         ['64', 'female', '32.965', '0', 'no', 'northwest', '14692.66935'],
         ['63', 'female', '26.98', '0', 'yes', 'northwest', '28950.4692'],
         ['60', 'female', '30.5', '0', 'no', 'southwest', '12638.195'],
         ['61', 'female', '25.08', '0', 'no', 'southeast', '24513.09126'],
         ['62', 'female', '39.2', '0', 'no', 'southwest', '13470.86'],
```

Figure 4: Displaying High-Risk Records Example

Those records are high-risk level records. Then, for further analysis, we only work on high-risk level records for this research.

E. D2R-Server Processing

D2R Server a tool for publishing the converted data into linked data on the web [19]. D2R server also provides functionality to convert the relational databases to linked data, so that it is ready to publish on the Semantic Web. The semantic web is a global information space which consists the data in Linked data format.

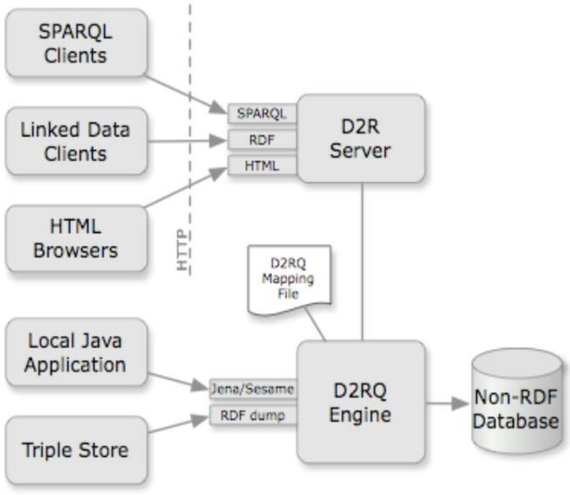


Fig 5: Publishing Data to Semantic Web

D2RQ engine provides functionality to query to the D2R engine. D2RQ engine supports most of the all traditional database and it allows query through SPARQL. The query will fetch the data into triples format. Below see the D2R server working for this work:

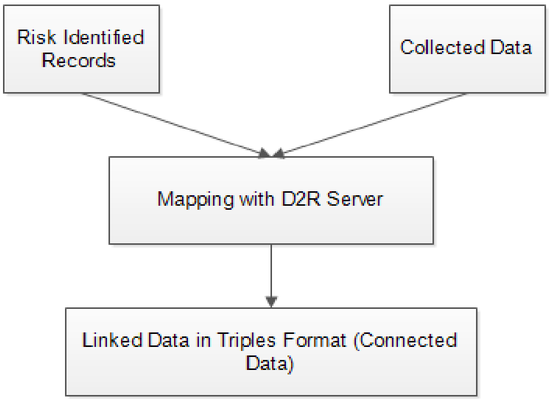


Figure 6: D2R Server Mapping of Data

In the mapping process both processed and unprocessed data mapped with insurer additional data. Therefore, advanced analysis completed through ontology which is mapped combined data.

F. Analyze High-Level Risk Records With Ontology

The ontology created with the mapping of the raw data and additional information of the insurer. Thus from the ontology, the insurance company person can fire query through SPARQL and get the whole analyzed information of the high-risk level insurer. With this advanced analytics, the insurance company can track particular user data by getting additional information from web sources. Thus, the risk will be minimized as additionally insurance company knows about insurer activity.

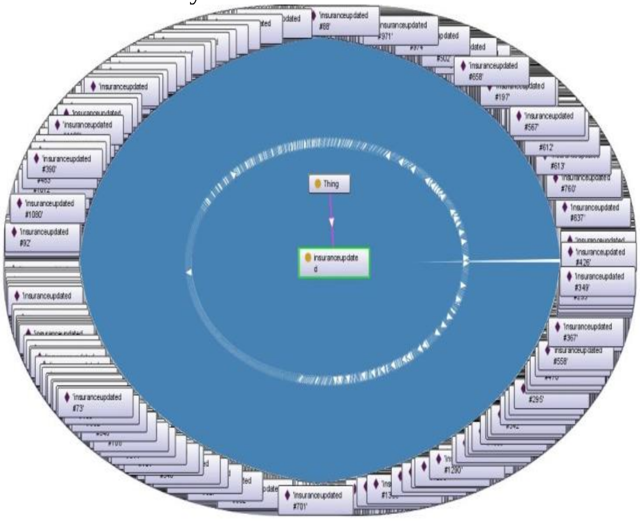


Fig 7: Mapped Ontology

After this results will be analyzed in cyclic days to know about the high-risk level insurers. Thus fraud detection can be identified, as data are fully mapped to the insurer and from the other sources also.

G. Classified Results

Now, we will see results and discussion for this study that analyze a score of advanced analysis and predict a certain number of people who are identified as high risk level category.

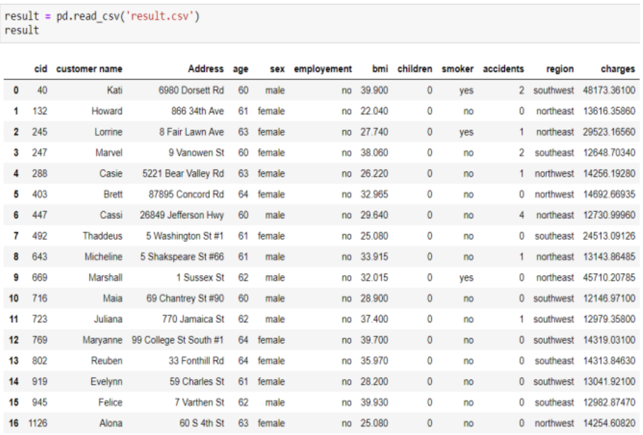
V. RESULTS AND DISCUSSIONS

In this study, we have developed a linked data approach to risk analysis in an insurance sector based on the structured data. Not to choose structured data and targeting linked data analysis in this study has a particular aim to achieve. In many industries or if we take life insurance industry the data collected often structured or stored in table formats. From that table or rows and columns, dynamic analysis cannot be proven beneficial. Thus linked data analysis dynamically analyzes the data which have relation to a particular person or entity.

The data processed in Python Jupyter notebook which provide an extensive environment for Python programming. Let see the particular code that produces the result:

Retrieval Number: J90430881019/19©BEIESP
DOI: 10.35940/ijitee.J9043.0881019
Journal Website: www.ijitee.org

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)



```
result = pd.read_csv('result.csv')
result
```

	cid	customer name	Address	age	sex	employement	bmi	children	smoker	accidents	region	charges
0	40	Kati	6980 Dorsett Rd	60	male	no	39.900	0	yes	2	southwest	48173.36100
1	132	Howard	866 34th Ave	61	female	no	22.040	0	no	0	northeast	13616.35860
2	245	Lorrine	8 Fair Lawn Ave	63	female	no	27.740	0	yes	1	northeast	29523.16560
3	247	Marvel	9 Vanowen St	60	female	no	38.060	0	no	2	southeast	12648.70340
4	288	Casie	5221 Bear Valley Rd	63	female	no	26.220	0	no	1	northwest	14256.19280
5	403	Brett	87895 Concord Rd	64	female	no	32.965	0	no	0	northwest	14692.66935
6	447	Cassi	26849 Jefferson Hwy	60	male	no	29.640	0	no	4	northeast	12730.99960
7	492	Thaddeus	5 Washington St #1	61	female	no	25.080	0	no	0	southeast	24513.09126
8	643	Micheline	5 Shakspeare St #66	61	male	no	33.915	0	no	1	northeast	13143.86485
9	669	Marshall	1 Sussex St	62	male	no	32.015	0	yes	0	northeast	45710.20785
10	716	Maia	69 Chantrey St #90	60	male	no	28.900	0	no	0	southwest	12146.97100
11	723	Juliana	770 Jamaica St	62	male	no	37.400	0	no	1	southwest	12979.35800
12	769	Maryanne	99 College St South #1	64	female	no	39.700	0	no	0	southwest	14319.03100
13	802	Reuben	33 Fonthill Rd	64	female	no	35.970	0	no	0	southeast	14313.84630
14	919	Evelynn	59 Charles St	61	female	no	28.200	0	no	0	southwest	13041.92100
15	945	Felice	7 Varthen St	62	male	no	39.930	0	no	0	southeast	12982.87470
16	1126	Alona	60 S 4th St	63	female	no	25.080	0	no	0	northwest	14254.60820

Fig 8: High-Level Risk Records Based on the Condition

The above figure displays the results based on the condition after generating rules. The high level risk values further converted to Linked data so that again geospatial information mapped to it and further analysis will be done on that.

Plot generation and analysis can be done through python Jupyter environment. The plotting and charting based on the external data and using internal data can be displayed. Below figure shows region wise higher risk pending charges.

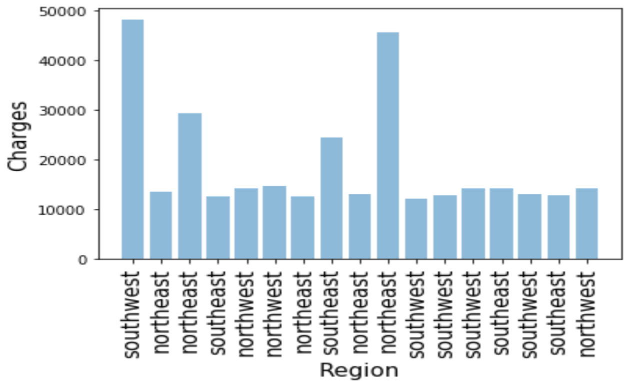


Fig 9: Region-wise Pending Charges

Ontology of the results data has all the attributes and is parsed to a link to be mapped with a web of data.

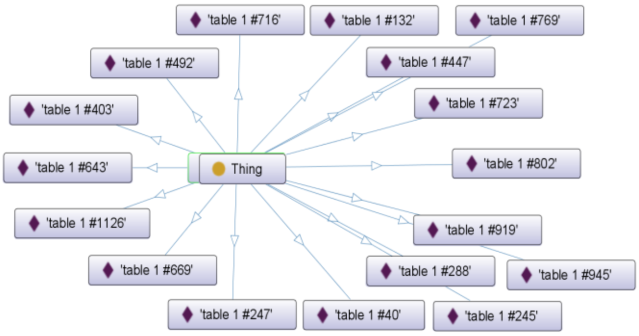
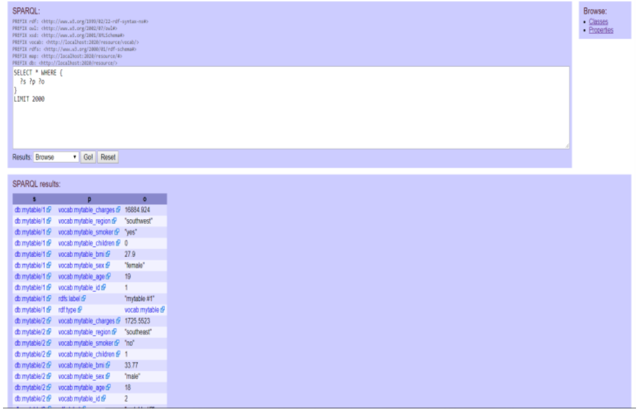


Fig 10: Ontology of Results

Thus the aim achieved after knowing that how easily we can do advanced analysis with the help of linked data. Linked data also placed easily on the Semantic web where machine understandable rich content acquired.



```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX vocab: <http://localhost:2020/resource/vocab/>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX map: <http://localhost:2020/resource/#>
PREFIX db: <http://localhost:2020/resource/>
SELECT * WHERE {
  ?s ?p ?o
}
LIMIT 2000
```

s	p	o
db:mytable/1	vocab:mytable_charges	16884.924
db:mytable/1	vocab:mytable_region	"southwest"
db:mytable/1	vocab:mytable_smoker	"yes"
db:mytable/1	vocab:mytable_children	0
db:mytable/1	vocab:mytable_bmi	27.9
db:mytable/1	vocab:mytable_sex	"female"
db:mytable/1	vocab:mytable_age	19
db:mytable/1	vocab:mytable_id	1
db:mytable/1	rdfs:label	"mytable #1"
db:mytable/1	rdf:type	vocab:mytable
db:mytable/2	vocab:mytable_charges	1725.5523
db:mytable/2	vocab:mytable_region	"southeast"
db:mytable/2	vocab:mytable_smoker	"no"
db:mytable/2	vocab:mytable_children	1
db:mytable/2	vocab:mytable_bmi	33.77
db:mytable/2	vocab:mytable_sex	"male"
db:mytable/2	vocab:mytable_age	18
db:mytable/2	vocab:mytable_id	2

Fig 11: SPARQL query execution o Linked Data Results

The above figure displays how data can be retrieved through SPARQL query language from Linked Data. The important part is that the SPARQL query fired on the knowledge base (ontology) finds appropriate RDF triples in Knowledge Base. It makes results precise when connected or linked data fetched in one instance.

		True Values	
	Total Values	*Positive High Risk Values*	*Negative High Risk Values*
Predicted Values	*Predicted condition positive*	16	1
	Predicted condition negative	9	24

(I)

For accuracy of the results, contingency matrix is created for results based on the sample data. The above Table 1 shows numeric terms that recorded after implementation. Sample data of 50 records from 1400 records taken for implementation from our database. It was found that 16 records are Positive True which gives 0.94 as precision value and 0.64 as recall value.

$$\textbf{Precision Value} = \frac{tp}{tp+fp} = \frac{16}{16+1} = 0.94 \qquad (2)$$

$$\textbf{Recall Value} = \frac{tp}{tp+fn} = \frac{16}{16+9} = 0.64 \qquad (3)$$

Where, *tp*, *fp* and *fn* are *True Positive*, *False Positive* and *False Negative* values respectively. And for the same we found the accuracy value as 0.8.

$$\textbf{Accuracy} = \frac{tp+tn}{tp+tn+fp+fn} = \frac{16+24}{16+24+1+9} = 0.80 \qquad (4)$$

Where *tn* is *True Negative* value. The implementation is performed on sample data in which our algorithm gives the following accuracy value. Therefore, with the following accuracy, the result data can be trusted for risk analysis for future events.

VI. CONCLUSION

The insurance companies in present days gives maximum efforts to identify fraudulent claims, insurer's risk, and areas that make a loss in the future. The loss generating in the life insurance industry make industry future perspective low. Thus the solution to map from the data or approach to analyze the data is changed now. This paper proposes linked data approach to identifying risk from insurer's data. The mapping of the data to the external sources data converts structured data to linked data. Linked data can be easily mapped with other data thus advanced analysis will be possible through it. The future scope of this study targets mapping of big data collected through web resources of user activity or through web scrapping to linked data.

REFERENCES

1. Risk Factors in Life Insurance https://www.finder.com/risk-factors-in-life-insurance
2. T. Berners-Lee. (2006, Jul.). Linked Data [Online]. Available: http://www.w3.org/designissues/linkeddata.html
3. C. Bizer, T. Heath, K. Idehen, and T. Berners-Lee, "Linked Data on The Web (Ldow2008)," In Proc. Www 2008, Pp. 1265–1266.
4. National Audit Office (Nao), "Staying the Course: The Retention of Students in Higher Education," London, 2007.
5. F. Sarker, H. Davis, and T. Tiropanis, "A Review of Higher Education Challenges and Data Infrastructure Responses," in Proceedings of International Conference for Education Research and Innovation (ICERI2010), Madrid, Spain, 2010.
6. V. Tinto, "Dropout From Higher Education: A Theoretical synthesis Of Recent Research," Review of Educational Research, Vol. 45, Pp. 89-125, 1975.
7. V. Tinto, "Leaving College: Rethinking The Causes And Cures Of Student Attrition", Contemporary Sociology, 17(3).
8. K. E. Arnold. "Signals: Applying Academic Analytics", Educause Quarterly (EQ) Magazine, 2010.
9. Farhana Sarkar et al., "Linked Data, Data Mining And External Open Data for Better Prediction of At-Risk Students", 2014 International Conference on Control, Decision and Information Technologies (CoDIT), DOI: 10.1109/CoDIT.2014.6996973.
10. V. Rawte And G. Anuradha, "Fraud Detection Using Data Mining Techniques," 2015, Vol. 4, No. 1, Pp. 304–312.
11. Framework For The Identification Of Fraudulent Health Insurance Claims Using Association Rule Mining
12. Dumbill, Edd. What Is Big Data? An Introduction To The Big Data Langscape. [Online] January 11, 2012. Http://Strata.Oreilly.Com/2012/01/What---Is---Big---Data.Html.
13. James Manyika, Michael Chui, Brad Borwn, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers. Big Data: The Next Frontier For Innovation, Competition, And Productivity. S.L.: Mckinsey Global Institute,2011. Http://Www.Mckinsey.Com/Insights/Mgi/Research/Technology_And_Innovation/Big_Data_The_Next_Frontier_For_Innovation.
14. Lishan Zhang, "How Structured Data (Linked Data) In Help in Big Data Analysis -- Expand Patent Data With Linked Data Cloud".
15. Jie Yuan et al., "A Linked Data Approach For Geospatial Data Provenance", IEEE Transactions on Geoscience and Remote Sensing, Vol. 51, Issue. 11, Pp. 5105-5112.
16. P. Yue, Y. Wei, L. Di, L. He, J. Gong, And L. Zhang, "Sharing Geospatial Provenance In A Service-Oriented Environment," Comput., Environ. Urban Syst., Vol. 35, No. 2, Pp. 333–343, 2011.
17. Raghu. (2018, April). Insurance, Version 1, Retrieved: August 20, 2019 From: https://www.kaggle.com/raghupalem/insurance
18. A. R. Cai Et Al., "Identification Of Adverse Interactions Through Causal Association Rule Discovery From Spontaneous Adverse Event Reports," Artif. Intell. Med.,2017.
19. D2R Server: A Semantic Web Front-End To Existing Relational Databases Richard Cyganiak And Christian Bizer Freie Universitat Berlin ¨ Richard@Cyganiak.De, Chris@Bizer.De

AUTHOR(S) PROFILE

Vishal Shah is pursuing his Master of Computer Applications in VIT university, Chennai. He has presented papers in reputed conferences and also has published research papers in Journals in the area of Cognitive computing and its applications. He has worked with the research projects of the VIT faculty. His research areas are Semantic-web, Software Engineering, Linked Data and Cognitive Computing applications.

Dr.Shridevi Subramanian is currently working in School of computing science and Engineering at VIT University. She completed her Doctorate from MS University. She has published many research papers in reputed International Journals. Her research interests are semantic technologies , web services and web mining.