DATA MINING PREPARATION: PROCESS, TECHNIQUES AND MAJOR ISSUES IN DATA ANALYSIS

Dr C K Gomathy-Assistant Professor, Department of CSE, SCSVMV Deemed to be University, India

Mr.P.Gopala Krishna Reddy, Mr.P Sai Srinivas, Mr.T.Karthikeya, Mr.R.Uday Kumar

UG Scholars, Department of CSE, SCSVMV Deemed to be University, India

Abstract:
Data preparation is an iterative-agile process for exploring, combining, cleaning and transforming raw data into curated datasets for self-service data integration, data science, data discovery, and BI/analytics. To perform data preparation, data preparation tools are used by analysts, citizen data scientists and data scientists for self-service. The tools are also used by citizen integrators and data engineers for data enablement to reduce the time and complexity of interactively accessing, cataloging, harmonizing, transforming and modeling data for analytics in an agile manner with metadata and lineage support. These tools can provide data access for use in mostly analytical tasks that include storage, logical and physical data modeling, and data manipulation for data visualization, data integration and analytics. Some tools support machine-learning algorithms that can recommend or even automate actions to augment and accelerate data preparation.

Keywords: Data Mining, Data preparation, Data Analysis, Big Data Analytics

1.Introduction:

Data preparation is the sorting, cleaning, and formatting of raw data so that it can be better used in business intelligence, analytics, and machine learning applications.

Data comes in many formats, but for the purpose of this guide we're going to focus on data preparation for the two most common types of data: numeric and textual.

Numeric data preparation is a common form of data standardization. A good example would be if you had customer data coming in and the percentages are being submitted as both percentages (70%, 95%) and decimal amounts (.7, .95) – smart data prep, much like a smart mathematician, would be able to tell that these numbers are expressing the same thing, and would standardize them to one format.

Textual data preparation addresses a number of grammatical and context-specific text inconsistencies so that large archives of text can be better tabulated and mined for useful insights.Text tends to be noisy as sentences, and the words they are made up of, vary with language, context and format (an email vs a chat log vs an online review). So, when preparing our text data, it is useful to 'clean' our text by removing repetitive words and standardizing meaning.For example, if you receive a text input of:'My vacuum's battery died earlier than I expected this Saturday morning A very basic text preparation algorithm would omit the unnecessary and repetitive words leaving you with: 'Vacuum's' [subject] died [active verb] earlier [problem] Saturday morning [time]' This stripped down sentence format is now primed to be much easier to be tabulated analytically – an AI analysis bot could now, for instance, easily calculate the number of text responses complaining of 'early' 'battery' -related failures. Furthermore, this could be routed to a relevant support team that can help the customer. Another easy application is the removal of text noise, which, when dealing with customer service, can take the form of URLs. URLs can trip up machine analysis because each is a unique word and near impossible to group. So, an intelligent AI text prep bot can be trained to recognize the format of a URL and be programmed to turn all instances in the simple, recognizable format of '--url--'. It's even more important for your system to use text analysis to detect user input variation. Just for starters, we want to make sure personal information that our users send us for support needs (phone numbers, email addresses) aren't included in our larger analytics approach – to ensure data privacy and to simplify our data.

For example, customers may write their address and apartment in the same line, versus two separate lines. If your data prep doesn't take this into account, you'll end up with a ton of confusing entries that will completely scramble results. Because of its detail-oriented nature and perceived heavy manpower requirements, data prep is menial work, but it's benefits are obvious. Monkeylearn's founder and CEO Raúl Garreta agrees that it can be tedious:

"I mean, it's not the most interesting thing to do, but if you are a data practitioner you find the importance of doing it and, as a result, you appreciate it," he said. Properly prepared, clean data, can make or break proper analysis and thus hamper or help entire business strategies. Let's get into the benefits of well-prepped data.

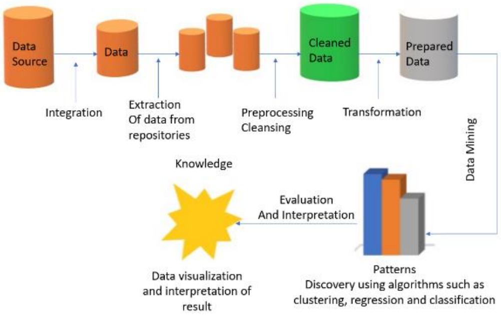
Data Source
Data
Cleaned Data
Prepared Data
Integration
Extraction Of data from repositories
Preprocessing Cleansing
Transformation
Data Mining
Knowledge
Evaluation And Interpretation
Data visualization and interpretation of result
Patterns Discovery using algorithms such as clustering, regression and classification

Figure 1 Process of Data Mining Preparation

Data mining is the central part of the method of knowledge detection. KDP is a process of seeking knowledge in data, using data mining methods (algorithms) to extract challenging knowledge from large quantities of information. The process of information exploration will consist of the following steps: 1- Data Source: In the context of computer science and computer applications, the source of the data is where the data used comes from. The source of data may be a database, a static file, live measurements on the physical system, abstract web data, or one of the myriad data services that stream and static over the internet. The primary concern for information accuracy is the data source. In such cases, to help companies and institutions operate more efficiently. Identifying data sources is the first step in any data storage project because you cannot do anything without the data. After setting up the right plan to obtain accurate information (data), the next step is to know how to store it consistently and in the same format so that when you run the reports, you can get the right results for decision making. Ultimately, data sources aim to help users and applications connect with and move data where it needs to be. 2- Data Cleaning: Data cleaning is the mechanism by which wrong, corrupt, misformatted, duplicate or incomplete data is corrected or deleted within a dataset. 3- Data Integration: Describes how data from multiple sources are integrated into usable and meaningful knowledge using a mix of techniques and business processes. 4- Data Selection: The term data selection seeks to choose data that should be preserved or shared/archived when the project is finished during the data collection. 5- Data Transformation: The transformation of data is the process of transferring data from one format or structure to another. For activities such as data integration and data management, data transformation is important. 6- Pattern Evaluation: The pattern evaluation describes fascinating patterns of information based on various types of interesting steps. A pattern is seen as appealing because it is potentially useful, readily understood by humans, and uses summary and viewing to make data understandable. 7- Knowledge Presentation: The representation of knowledge is a show of knowledge to the user in terms of trees, tables, rules, graphs, charts, matrices, etc. Represented one Methods for understanding the ins and outs of preparing data.

analysis should be submitted and therefore the data should be converted to fulfill these demands. Moreover, the selection of specific data to be analyzed can greatly influence the models learned. It is often the most time-consuming part of any data mining project. Many researchers have recognized high-end data extraction expertise and information as essential. research subject in machine learning and the database system and in many industrial enterprises as an interesting field with the potential of generating substantial revenues. Data, information, or knowledge have an important role in human activities. Data mining has significance in finding patterns, forecasting and discovering knowledge, etc. in various business fields. Data mining is used in the Medical Sciences, the detection of malicious executables, statistical techniques, identifying patterns, sales forecasting, basket analysis, mathematical organizations, and the presentation of information in a way that can be easily interpreted by people. This allows businesses and organizations to concentrate their stored data on the most relevant information.The massive increase in data in recent years has led to this, prompting recording, processing, and analyzing these records. Research Contributions To automate data preparation, several challenges must be addressed, including: (1) accommodating several different components of the same task, Establish alternative approaches; (2) Coordination of the various components of data preparation that rely on different evidence about the problem to be solved; (3) Identify the evidence that enables us to perform steps on the data; (Iv) Creating several options between candidate and alternative solutions, in light of multiple requirements; (5) Linking multiple components for the purpose of forming transactions.

2.Literature Survey and Proposed System:

Data mining is important in pattern finding, prediction, discovery of knowledge, etc. in different fields of industry. Data mining applications use a range of data types, from text to photos, warehouses, and different databases and data structures. Different data mining techniques for extracting patterns and hence knowledge from these various databases. Data collection and methods Data mining is an essential activity and domain awareness is crucial in this process. A range of data needs to be collected in the particular problem area to collect data, pick data from the identified data for data mining, clean and process data, extract patterns to generate inforsmation and finally interpret pattern and generate knowledge. Data mining is used in medical sciences, malicious executables tracking, sports associations, trend recognition, sales forecasting, basket analysis etc. There were still numerous unresolved security problems, social concerns, user interface issues, performance issues, and so on before data mining became a conventional, mature and trusted file. While data mining is very efficient, during its implementation it faces numerous challenges. Problems and challenges related to data mining may be efficiency, data, techniques, etc. When challenges or problems are correctly identified and sorted properly, data mining is successful. We would like to propose potential recommendations for data creation, including the development of efficient and successful data prepared algorithms and systems for single and multiple data sources, taking all internal data into account. And external knowledge. Create the environment for immersive and automated data extraction.

3. Methodology:

Data mining is the process by which useful data, patterns and trends from several data are collected, using techniques like clustering, classification, regression and correlation. Data, information, and knowledge are the exciting roles of human life. Massive data warehouses with the rapid development of file technologies require big data analysis and modelling to predict future information trends. Data mining is called a technique by which the necessary information in databases can be extracted from raw information. Using the data mining prediction analysis methodology, future scenarios can be predicted with regard to current knowledge. Forecast analysis is a combination of classification and aggregation. Data mining is used for data extraction from a great deal of information. Data mining is made up of two predictive and descriptive models. Data management aims to collect data in stratification files with either the ultimate aim of learning new effects or seeing new areas. Data analysis techniques were applied to the higher education institution's educational data. The analytical data included event records which extracted hidden information from the data by performing pattern recognition and forecast modelling tasks. Solve many complex problems, including energy efficiency and energy use, structural analysis, building materials, smart cities, design and optimization, technology forecasts, soil engineering and construction engineering. Accurate forecasting of traffic information, predicting the occurrence of COVID-19, the user must use the information published effectively through prediction models. In this section, the applied methods used in our approaches, such as decision support systems, data mining, and correct data preparation and extraction, have an effective role in showing the right results through their application to data mining techniques. The results differed between accurate and imprecise due to their reliance on the techniques and algorithms used to show results on prepared data. Therefore, the main features of the conceptual data management platform must be followed to prepare the data to offer more accurate results before implementing it, which are:

1. Taking the initial data and preparing it for analysis.
2. Apply algorithms to raw data to reveal new insights.
3. Create database categories to place raw data.
4. Data collection and classification.
5. Collect data from multiple sources and collect them together.
6. Converting unstructured data into data ready for analysis.
7. Merging different types of data into a unified system.
8. Converting data from one type to data of another kind.

2. Data Mining Implementation Process Many different sectors are leveraging data mining to enhance their businesses' efficiency, including manufacturing, banking, marketing, aerospace, education, health, etc. Therefore, the need for the traditional data extraction process has effectively improved. Data mining techniques should be reliable and reproducible by company personnel with little or no knowledge of the data mining context. Several steps determine the classic presentation (see Table 3). First, The problem has to be identified in terms of work or academic objectives and converted into concrete data mining and analytical objectives. The second stage is the data step, the discovery, fusion and transformation of primary data sources to be used for the related data mining mission. This is typically the longest step unless the process is completely automated. The third step is step modelling. Algorithms are used to extract real data patterns, to predict or to calculate my metadata. In the fourth stage these patterns and models are evaluated in a quality and content format. The extracted forms will be added to the new data during the final publication stage and the findings are combined with other details for effective action.

3. Major Issues in Data Mining With the increasing growth of data in any application, data mining satisfies an imminent need for accurate, scalable, and flexible data analysis in our society Data mining can be seen as a natural IT creation and a convergence of several disciplines and associated fields of application. Although data mining is very effective, it faces numerous challenges throughout its implementation. Data mining issues and challenges can involve performance, data, techniques, etc. Data mining is successful if issues or problems are correctly detected and properly sorted. Data can be collected on any data as long as data is essential for the target app, such as database data, data warehouse data, transaction information, and advanced data forms. Figure 2 demonstrates the data mining problems.

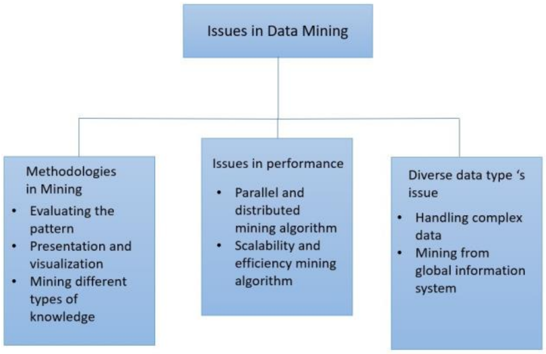
Issues in Data Mining
Methodologies in Mining
• Evaluating the pattern
• Presentation and visualization
• Mining different types of knowledge
Issues in performance
• Parallel and distributed mining algorithm
• Scalability and efficiency mining algorithm
Diverse data type 's issue
• Handling complex data
• Mining from global information system

Figure 2 Issues in data mining

4. Data mining techniques Organizations and institutions today have more access than ever to data. However, it can be very difficult to grasp large volumes of structured and unstructured data to strengthen the organization and at other levels, due to the sheer volume of information. If not properly treated, this challenge will lessen the benefits of all data. Data mining is how businesses discover data trends to obtain knowledge that is important to their business needs. For both business intelligence and data science, it is important. There are many techniques for data mining that organizations can use to convert raw information into actionable insights. This encompasses everything from advanced artificial intelligence to fundamental data planning, which are the secret to optimizing the value of data investment.

4.1. Classification

It's a task of data analysis, i.e. the process of finding a model that describes and differentiates data

classes and concepts. This data extraction approach helps to classify data into different categories.

4.2. Clustering

Cluster analysis is a form of data extraction to classify related data. This method helps to consider
gaps between data and similarities.

4.3. Regression

Regression analysis is a data extraction process in which the relationship between variables is defined
and analyzed. It is used to evaluate a given variable 's likelihood, since there are other variables.

4.4. Association rules

This data extraction technique helps to find the connection between two or more objects. Detects a
pattern in the dataset.

4.5. Outer detection

This type of data extraction technique refers to the observation of data elements that do not fit the
predicted behavior pattern in the data collection. This technology can be used in different areas
including intrusion, tracking, fraud, detection of bugs, etc. Offshore analysis or offshore mining is also
known as external detection.

4.6. Sequential patterns

This technique of data collection helps to detect or discover similar patterns or trends in transaction
data over a given timeframe.

4.7. Prediction

The predictions used numerous other techniques for data mining, such as patterns, sequences,
clustering, grouping, etc. It analyses past events or circumstances in the right order to predict a future
occurrence.

4 .Implementation & Algorithm Techniques:

Access

There are many sources of business data within any organization. Examples include endpoint data, customer data, marketing data, and all their associated repositories. This first essential data preparation step involves identifying the necessary data and its repositories. This is not simply identifying all possible data sources and repositories, but identifying all that are applicable to the desired analysis. This means that there must first be a plan that includes the specific questions to be answered by the data analysis.

Ingest

Once the data is identified, it needs to be brought into the analysis tools. The data will likely be some combination of structured and semi-structured data in different types of repositories. Importing it all into a common repository is necessary for the subsequent steps in the pipeline. Access and ingest tend to be manual processes with significant variations in exactly what needs to be done. Both data preparation steps require a combination of business and IT expertise and are therefore best done by a small team. This step is also the first opportunity for data validation.

Cleanse

Cleansing the data ensures that the data set can provide valid answers when the data is analyzed. This step could be done manually for small data sets but requires automation for most realistically sized data sets. There are software tools available for this processing. If custom processing is needed, many data engineers rely on applications coded in Python. There are many different problems possible with the ingested data. There could be missing values, out-of-range values, nulls, and whitespaces that obfuscate values, as well as outlier values that could skew analysis results. Outliers are particularly challenging when they are the result of combining two or more variables in the data set. Data engineers need to plan carefully for how they are going to cleanse their data.

Format

Once the data set has been cleansed; it needs to be formatted. This step includes resolving issues like multiple date formats in the data or inconsistent abbreviations. It is also possible that some data variables are not needed for the analysis and should therefore be deleted from the analysis data set. This is another data preparation step that will benefit from automation. Cleansing and formatting steps should be saved into a repeatable recipe data scientists or engineers can apply to similar data sets in the future. For example, a monthly analysis of sales and support data would likely have the same sources that need the same cleansing and formatting steps each month.

Combine

When the data set has been cleansed and formatted, it may be transformed by merging, splitting, or joining the input sets. Once the combining step is complete, the data is ready to be moved to the data warehouse staging area. Once data is loaded into the staging area, there is a second opportunity for validation.

Analyze

Once the analysis has begun, changes to the data set should only be made with careful consideration. During analysis, algorithms are often adjusted and compared to other results. Changes to the data can skew analysis results and make it impossible to determine whether the different results are caused by changes to the data or the algorithms.

5.CONCLUSION:

In this chapter, we have discussed several methods and techniques for data preprocessing in the
context of Big Data. We have reported experiences and first insights about the preprocessing of a
real world dataset in a petro-chemical production setting. Overall, the principle “privacy by design” is extremely important in a big data environment in order to protect personal data of individuals from further analytics. We also identified two types of tasks that have to be addressed separately in order to create an assistant system for early warnings. Our experiences show that unstructured data can be found in various places in a production environment containing shift reports, alarm data and even some error codes in the sensor data. For structured data, always the relation between filtering and information loss needs to be balanced. Furthermore, one simple, but important preprocessing techniques for the analysis of natural language text is the mapping of different notations of the same word to a common form, i.e., to a common terminology so that the different entities can be correctly resolved. Furthermore, we have found that the conversion of file formats can lead to further preprocessing steps, especially when the data is fragmented. Finally, it is important to consider that data is not always coming from one system, making it necessary to check for time offsets and the reasons
behind it, connecting that to the business and data understanding phases of CRISP-DM.

6.Results:

One key aspect prior to an analysis is the anonymization of the data. In order to protect personal data of individuals from further analytics, person names should be made unrecognizable. One way to achieve this is to simply remove those names from the document. This follows the principle "privacy by design" which means that anonymization should be performed as early as possible, so that person names cannot be the subject of further analytics any more. When it comes to automated anonymization, it is also a requirement to convert the files from binary to text formats. One should take into consideration that such a conversion could cause a loss of information. For example, in the case of graphical documents, it is obvious that the visual information cannot be captured by a text format. Such documents have to be manually anonymized by hand (e.g., by blackening person names). For text processing documents (e.g., word format) most of the information can be preserved by choosing an html format over a plain text format. This way, the document structure (e.g., headlines, bold words) is still available for analytical processing such as generating warnings for abnormal situations. An assistant system that generates early warnings for abnormal situations actually has to solve two types of tasks. At first, the system has to identify events based on the long history of data. A burst in the frequency of the alarm logs, for example, could be an indicator for an unexpected situation, which corresponds to an event. Secondly, after identifying events, the system needs to extract features that help to predict this type of event as early as possible. Coming back to the frequency of alarm logs example, small fluctuations in the frequency distribution could be an indicator for a specific event characterized by a burst in the alarm log frequency. In our analysis, we focused on unstructured data, i.e., free text entered by operators into the system, but we found that text data could also be part of the sensor data as well. For example, a sensor value is only recorded when it is within a specific range of electricity current, e.g. 24mA. An electricity current above the threshold cannot be recorded and will result in a specific error code translated to a label, e.g. the character string "bad value". There exists a variety of error codes for sensor data and one has to consider how to deal with these error values. When analyzing free text entered into the system by an operator one has also to consider two types of situations. Text messages that are really typed by the operator, e.g. because the text is about irregular situation in the production facility, and text messages that are just copied from a template, e.g. the text is about a standard procedure like weekly maintenance work. Both situations show different characteristics and should be considered separately. In the former case, the free text is prone to typos and different spellings of the words, e.g. abbreviations and acronyms. In an industrial environment, there also exist some special wordings and a domainspecific vocabulary that has to be taken into account. One of the key challenges when working with industrial text data is to define a semantical relatedness between the words due to the lack of domain-specific concept hierarchies. In the FEE project, we tried to overcome this problem by using two approaches. First, we used stemming (see section "Preprocessing of Unstructured Data") in order to find semantically related words by removing of prefixes and suffixes. Furthermore, we used a micro-worker approach, where ordinary persons should mark similar words in order to improve the set of related words. We found that the micro-worker approach was difficult due to the lack of domainspecific knowledge by the micro-workers. With stemming, there is also the problem of over- and under-stemming making it necessary to manually inspect stemming results. Concerning numerical data, methods like outlier detection can help to check the data quality and to ensure meaningful episodes, e.g., in the case of time series. In addition, value imputation methods can be helpful, e.g., in the case of missing values. Advanced methods for data preprocessing, like filter approaches of course can result in a loss of information and should therefore always be targeted with respect to the analytical goals. Having different data sources in FEE project, we also had to deal with different file formats. All the binary formats were converted to text formats in order to be able to further process the files with standard command line tools. For Excel documents, we choose the CSV format and Word / PDF documents were converted to HTML format in order to keep as much information about the formatting as possible. With the PDF documents, we found that text spanning multiple lines is divided by line when converted to a text format. This could lead to artefacts that have to be further processed, e.g. by dense-based clustering, to connect associated character strings again. We also like point out, that when dealing with data from different data sources one has to consider the time dependency of the data. In a production environment, there are multiple production applications that have been introduced and expanded over multiple years. These systems are typically not synchronized by a local time server making it necessary to inspect time offsets between different data sources. We found that log entries in the operation journals had an offset of approximately one hour when compared to events in the sensor data (see Figure 2). Reasons for time offsets can be found in differing time zones as well as data types not capable of dealing with daylight saving time. Furthermore operators do not have time for documentation when a critical situation is about to happen, because they have to react to bring the system to a stable state again. Therefore, most of the documentation for critical situations is done after the event has happened.

7.References:

1.DR.C.K.Gomathy , V.Geetha , S.Madhumitha , S.Sangeetha , R.Vishnupriya Article: A Secure With Efficient Data Transaction In Cloud Service, Published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 4, March 2016, ISSN: 2278 – 1323.

2.Dr.C.K.Gomathy,C K Hemalatha, Article: A Study On Employee Safety And Health Management International Research Journal Of Engineering And Technology (Irjet)- Volume: 08 Issue: 04 | Apr 2021

3. Dr.C K Gomathy, Article: A Study on the Effect of Digital Literacy and information Management, IAETSD Journal For Advanced Research In Applied Sciences, Volume 7 Issue 3, P.No-51-57, ISSN NO: 2279-543X,Mar/2018

4. Dr.C K Gomathy, Article: An Effective Innovation Technology In Enhancing Teaching And Learning Of Knowledge Using Ict Methods, International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) E-Issn: 2395-5325 Volume3, Issue 4,P.No-10-13, April '2017

5.Dr.C K Gomathy, Article: Supply chain-Impact of importance and Technology in Software Release Management, International Journal of Scientific Research in Computer Science Engineering and

Information Technology (IJSRCSEIT) Volume 3 | Issue 6 | ISSN : 2456-3307, P.No:1-4, July-2018.

6.C K Gomathy and V Geetha. Article: A Real Time Analysis of Service based using Mobile Phone Controlled Vehicle using DTMF for Accident Prevention. International Journal of Computer Applications 138(2):11-13, March 2016. Published by Foundation of Computer Science (FCS), NY, USA,ISSN No: 0975-8887

7.C K Gomathy and V Geetha. Article: Evaluation on Ethernet based Passive Optical Network Service Enhancement through Splitting of Architecture. International Journal of Computer Applications 138(2):14-17, March 2016. Published by Foundation of Computer Science (FCS), NY, USA, ISSN No: 0975-8887

8.C.K.Gomathy and Dr.S.Rajalakshmi.(2014), "A Software Design Pattern for Bank Service Oriented Architecture", International Journal of Advanced Research in Computer Engineering and Technology(IJARCET), Volume 3,Issue IV, April 2014,P.No:1302-1306, ,ISSN:2278-1323.

9.C. K. Gomathy and S. Rajalakshmi, "A software quality metric performance of professional management in service oriented architecture," Second International Conference on Current Trends in Engineering and Technology - ICCTET 2014, 2014, pp. 41-47, doi: 10.1109/ICCTET.2014.6966260.

[11] Dr.C K Gomathy, V Geetha ,T N V Siddartha, M Sandeep , B Srinivasa Srujay Article: Web Service Composition In A Digitalized Health Care Environment For Effective Communications, Published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 4, April 2016, ISSN: 2278 – 1323.

[12]C.K.Gomathy.(2010),"Cloud Computing: Business Management for Effective Service Oriented Architecture" International Journal of Power Control Signal and Computation (IJPCSC), Volume 1, Issue IV, Oct - Dec 2010, P.No:22-27, ISSN: 0976-268X .

[13]Dr.C K Gomathy, Article: A Study on the recent Advancements in Online Surveying , International Journal of Emerging technologies and Innovative Research (JETIR) Volume 5 | Issue 11 | ISSN : 2349-5162, P.No:327-331, Nov-2018

[14]Dr.C.K.Gomathy,C K Hemalatha, Article: A Study On Employee Safety And Health Management International Research Journal Of Engineering And Technology (Irjet)- Volume: 08 Issue: 04 | Apr 2021

[15] Dr.C K Gomathy, V Geetha , T.Jayanthi, M.Bhargavi, P.Sai Haritha Article: A Medical Information Security Using Cryptosystem For Wireless Sensor Networks, International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) E-Issn: 2395-5325 Volume3, Issue 4, P.No-1-5,April '2017

[16] C.K.Gomathy and Dr.S.Rajalakshmi.(2014), "Service Oriented Architecture to improve Quality of Software System in Public Sector Organization with Improved Progress Ability", Proceedings of ERCICA-2014, organized by Nitte Meenakshi Institute of Technology, Bangalore. Archived in Elsevier Xplore Digital Library, August 2014, ISBN:978-9-3510-7216-4.

[17] Parameshwari, R. & Gomathy, C K. (2015). A Novel Approach to Identify Sullied Terms in Service Level Agreement. International Journal of Computer Applications. 115. 16-20. 10.5120/20163-2253.

[18] C.K.Gomathy and Dr.S.Rajalakshmi.(2014),"A Software Quality Metric Performance of Professional Management in Service Oriented Architecture", Proceedings of ICCTET'14, organized by Akshaya College of Engineering, Coimbatore. Archived in IEEE Xplore Digital Library, July 2014,ISBN:978-1-4799-7986-8.

[19] C.K.Gomathy and Dr.S.Rajalakshmi.(2011), "Business Process Development In Service Oriented Architecture", International Journal of Research in Computer Application and Management (IJRCM) ,Volume 1,Issue IV, August 2011,P.No:50-53,ISSN : 2231-1009

8. Authors Profiles

1.P.Gopala Krishna Reddy Student , B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. His Area of Interest in Big Data Analytics.

2.P.Sai Srinivas Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. His Area of Interest in Big Data Analytics.

3.T.Karthikeya Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. His Area of Interest in Big Data Analytics.

4.R.Uday Kumar Student, B.E. Computer Science and Engineering, Sri Chandrasekharendra SaraswathiViswa Mahavidyalaya deemed to be university, Enathur, Kanchipuram, India. His Area of Interest in Big Data Analytics.
5. Dr. C.K. Gomathy is Assistant Professor in Computer Science and Engineering at Sri Chandrasekharendra Saraswathi Viswa Mahavidyala, Enathur, Kanchipuram, India. Her area of interest: Software Engineering, Web Services, Knowledge Management and IoT