

# **RESEARCH PROPOSAL ON DATA SCIENCE AND ANALYTICS**

## **1. Introduction**

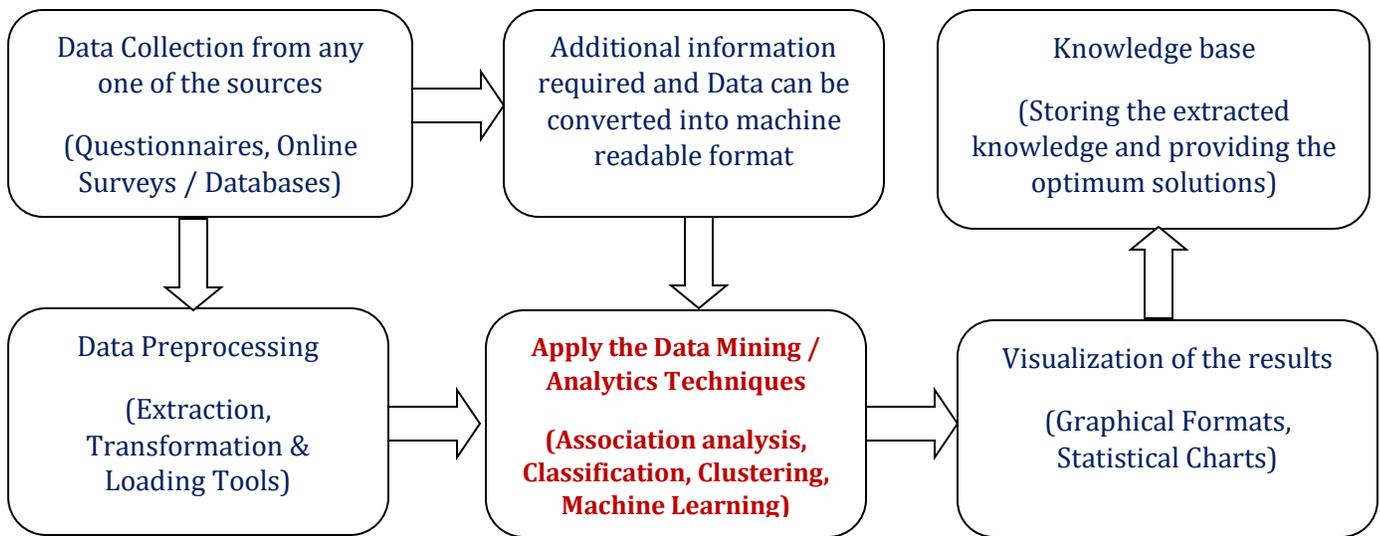
Data Science has been established as an important emergent scientific field and paradigm driving research evolution in such disciplines as statistics, computing science and intelligence science, and practical transformation in such domains as science, engineering, the public sector, business, social science, and lifestyle. The field encompasses the larger areas of artificial intelligence, data analytics, machine learning, pattern recognition, natural language understanding, and big data manipulation. It also tackles related new scientific challenges, ranging from data capture, creation, storage, retrieval, sharing, analysis, optimization, and visualization, to integrative analysis across heterogeneous and interdependent complex resources for better decision-making, collaboration, and, ultimately, value creation. We are living in the age of data, where nearly every task we conduct in our daily lives depends on data and can be tracked and supported digitally. Massive data of different types, including numeric variables, images, videos, music, text, etc., could be collected when shopping, working, socializing, communicating, relaxing and traveling, as part of our daily lives. As a multi-disciplinary field that integrates mathematics, statistics and computer science, data science uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data, with the ultimate goal to support decision making.

Hence, I have planned to do my research in the context of Data Science and Analytics. Data will be retrieved from online databases and developing a generic model for providing the optimum solutions for the chosen problem statement.

## **2. Aim and Methodology**

The primary objective of this research proposal is to select the appropriate problem statement along with the required datasets and to develop the model to provide the optimum results.

The entire workflow of this research work is illustrated in the figure1.



**Figure-1 Workflow of the proposed research work**

The workflow of the proposed research work can be divided into two parts viz. Data Preprocessing and Post processing. Data Preprocessing involves the data collection from anyone of the sources which may be in online surveys, forms, databases or questionnaires. Add the additional information for the collected input data. The next step is to apply any one of the Extraction, Transformation and Loading (ETL) tools. In this preprocessing, automatically the proposed research methodology will include the required techniques to have the ability to manage semi-structured or unstructured data.

After the completion of preprocessing, advanced analytics techniques (data mining or machine learning techniques) will be applied to the dataset. The main aim of the analytics is to predict the future by using the analysis of past and present data. When we applied these techniques, the performance measures will take part of this research work. Once, applying these measures, the proposed research model will show its robustness and efficiency of the research work.

In addition to these, the proposed research work includes the visualization techniques as well as maintaining the knowledge base for storing the extracted meaningful information. This knowledge base will be the best optimum solution for the new problem statements with the same kind of attributes.

### 3. Conclusion

Data are one of the most important assets during a crisis. Therefore, the proposed research work is used to apply the data science and analytics on any kind of data. From this research work, I have planned to develop a generic data analyzer tool to apply the advanced analytical tools such as data mining techniques (association analysis, classification, clustering) or machine learning algorithms. By applying these analytical techniques, we can provide the best optimum results for any type of research problem statements.

### 4. Journal Published:

- ✓ R. Syed Ali Fathima, "**Virtual Reality in Education and Training**", In the International Journal of Research in Advent Technology (IJRAT), May 2019, Vol.7 No.5S, pp. 204-207, May 2019.
- ✓ R. Syed Ali Fathima, "**A Study on Virtual Reality**", In the Shanlax International Journal of Arts, Science and Humanities and International Conference on Virtual Reality Techniques – ICVRT'18, PG Departments of Computer Applications & IT, Fatima College, Madurai, Vol 6, Special Issue 1, pp:103-108, 2018.
- ✓ R. Syed Ali Fathima, M.John Basha and P.Saravanan, "**Improving Security and Efficiency in Association Rule Mining using PFP-Growth Algorithm via Transaction Splitting**", In the SSRG International Journal of Computer Science and Engineering (SSRG-IJCSE) , vol 3 Issue 4, April 2016.
- ✓ R. Syed Ali Fathima and Dr.R.Lawrance, "**Surveillance Of Association Rule Mining**", In the ANJAC Journal of Science , Vol 12, no 2, August 2013.
- ✓ R. Syed Ali Fathima and Dr.R.Lawrance, "**A New Hybrid Search Based Algorithm Using Partition-COFI Tree – In Association Rule Mining**", In the CIIT International Journal of Data Mining and knowledge Engineering , Vol 4, no 7, July 2012.

## 5. List of Paper Presentation

- ✓ R. Syed Ali Fathima, “**Text Mining: A Brief Introduction**”, in the National Conference on Business Analytics, PG Department of Computer Applications & Management Studies, PG Department of Computer Applications & Management Studies, Fatima College(Autonomous), Madurai, pp:134-138, Sep 2019.
- ✓ R. Syed Ali Fathima, “**A Survey On Big Data Analytics And Tools**”, in the National Conference on Business Analytics, PG Department of Computer Applications & Management Studies, PG Department of Computer Applications & Management Studies, Fatima College(Autonomous), Madurai, pp:134-138, Sep 2019.
- ✓ R. Syed Ali Fathima, “**Virtual Reality in Education and Training**”, In the National Conference on Smart Innovative Technologies on Data Analytics, Department of Computer Applications, ANJAC, pp:141-145, Apr 2019.
- ✓ R. Syed Ali Fathima, “**A Study on Virtual Reality**”, In the International Conference on Virtual Reality Techniques – ICVRT’18, PG Departments of Computer Applications & IT, Fatima College, Madurai, Vol 6, Special Issue 1, pp:103-108, 2018.
- ✓ R. Syed Ali Fathima “**Comparison of FP Growth and PFP growth algorithm in Association Rule Mining via Transaction Splitting**”, in the proceedings of International Conference on Global Talent Management in the Digital Era, organized by the department of MBA, MCA, and M.Sc. IT, Fatima College(Autonomous), Madurai, on September 13, 2017.
- ✓ R. Syed Ali Fathima and Dr.R.Lawrance, “**Comparison of Association Rule Mining**”, In the proceedings of National Conference on Computer Science and Applications (NCCSA14), Mepco Schlenk Engineering college, Sivakasi, March 5,2014.
- ✓ R. Syed Ali Fathima and Dr.R.Lawrance, “**A New PC tree Algorithm for finding Frequent Items Using Multilevel Association Rules**”, In the proceedings of National Seminar on Advances in Computing techniques (NSACT 2012), SFR College for women, Sivakasi, September 25 & 26,2012.
- ✓ R. Syed Ali Fathima and Dr.R.Lawrance, “**Hybrid Search Based Algorithm for Finding Frequent Pattern Mining Using PC Tree – In Association Rule Mining**”, In the proceedings of International Conference on Mathematical Modeling and Applied Soft

Computing (MMASC 2012), Coimbatore Institute of Technology , Coimbatore, July 11-13,2012.

✓ R. Syed Ali Fathima, S. Sumathi and Dr.R.Lawrance, “**A Frequent Pattern Mining Algorithm Using Multiple Longest Common Subsequences (MLCS)**”, In the proceedings of International Conference of Recent Trends in Computer Science and Engg.(ICRTCSE 2012), Apollo Engineering College, Kanchipuram, May 3 & 4,2012.

✓ R. Syed Ali Fathima and Dr.R.Lawrance, “**A New Hybrid Search Based Algorithm Using Partition-COFI Tree - In Association Rule Mining**”, In the proceedings of International Conference on Computing and Control Engineering (ICCCE 2012), Dr.M.G.R. Educational and Research Institute University, Chennai, April 12 & 13,2012.

✓ R. Syed Ali Fathima and Dr.R.Lawrance , “**An effective Association Rule Mining in Dataset Using FuzzyLCS**”, In the Proceedings of International Conference on Recent Trends in Information Processing and Computing (ICIPC 2012), Nadar Saraswathi College of Arts and Science, Theni, January 6 & 7, 2012.

✓ R. Syed Ali Fathima, K. Poornamala and Dr.R.Lawrance, “**An Apriori-like algorithm for Extracting Fuzzy Association Rules between Keyphrases in Text Documents**”, In the Proceedings of National Conference on UGC Sponsored Coding Theory and Cryptography, Fatima College (Autonomous), Madurai, 07.09.2011 and 08.09.2011.