

Biosociolink: A Decision Support System for Analyzing Link Properties in Biological and Social Networks



Amulyashree Sridhar, Sharvani GS, Ramakanth Kumar P, AH Manjunatha Reddy, Kalyan Nagaraj

Abstract: Network based data representation has received widespread attention over the years. Data is oriented in graph format by aligning information as nodes and edges. Some of the predominant network cases include biological and social sciences. There is a growing need to extract knowledge patterns from network orientations. In such scenario, the current study focuses on extracting data patterns from network data. Schizophrenia gene data and TRAI wireless performance data is identified for performing biological and social network analysis. Biological network analysis is performed to identify relevant gene ties which act as hotspots for identifying disease causing genes. On similar lines, social network analysis is performed on wireless dataset to identify essential telecom operators responsible for customer retention. Based on these outcomes, a decision support system, BioSocioLink is designed in R programming language to perform biological and social network analysis. The support system accurately detects knowledge patterns from both the datasets. The study is concluded by deploying the support system in local programming environment.

Keywords: Network theory, biological network, social network, decision support system

I. INTRODUCTION

Networks represent structural and behavioral orientations of complex systems in real time. These entities are usually represented as a graph comprising of objects as nodes and their interactions as links [1]. Recent growth of data has resulted in generating mammoth information. However,

traditional database systems fail to store such data as tabular transactions. In this context, storing such information as nodes and links helps in preserving the semantic information [2, 3]. Over the past decade several network based studies have been proposed for addressing various problems including node classification [4], link prediction [5] and community detection [6]. Some of the popular networks include biological and social networks.

Biological networks represent biological processes captured by RNAs, genes and proteins. Each of these entities signifies different biological networks where nodes represent biological objects while links indicate the interconnectivity among them [7]. Some of the prominent biological networks include metabolic networks, gene interaction networks, protein-protein interaction networks, cell based networks and transcriptional networks [8]. Information captured from such networks helps in ascertaining interconnections across phenotype and genotype for any dreadful disease. Furthermore, several other network properties can be captured from biological networks including link connectivity, structural and functional orientations.

Social networks are other instance of network representations where objects are categorized as nodes and their interactions [9]. With exponential digital growth in the recent years, social networks have gained proficient importance in web page categorization, spam detection, ego discovery and recognition of web communities [10]. Online social networking sites like FaceBook, Twitter and LinkedIn have empowered congregation of mammoth data. Advanced algorithms are employed on such information for detecting internal dynamics and behavioral correlations among social entities. Henceforth, network based approach is found to be appropriate for managing biological and social data. Furthermore, decision making needs to be precise in such analysis to support accurate information processing. In this context, development of decision support system helps in generating accurate decisions [11]. A decision support system (DSS) aids in decision making process by analyzing network data to demonstrate real time information patterns. Based on these inputs, the current study focuses on development of decision support system for biological and social networking data to disclose valuable interactions among these entities. The two fold decision support system is developed in R programming language for enhancing decision making process among networks.

Revised Manuscript Received on October 30, 2019.

* Correspondence Author

Amulyashree Sridhar*, Department of Computer Science & Engineering, RV College of Engineering, Bangalore, affiliated to Visvesvaraya Technological University, Belagavi, India. Email: 0908amulyashree@gmail.com

Sharvani GS, Department of Computer Science & Engineering, RV College of Engineering, Bangalore, affiliated to Visvesvaraya Technological University, Belagavi, India. Email: sharvanigs@rvce.edu.in

Ramakanth Kumar P, Department of Computer Science & Engineering, RV College of Engineering, Bangalore, affiliated to Visvesvaraya Technological University, Belagavi, India. Email: ramakanthkp@rvce.edu.in

AH Manjunatha Reddy, Department of Biotechnology, RV College of Engineering, Bangalore, affiliated to Visvesvaraya Technological University, Belagavi, India. Email: ahmreddy@gmail.com

Kalyan Nagaraj, Department of Computer Science & Engineering, RV College of Engineering, Bangalore, affiliated to Visvesvaraya Technological University, Belagavi, India. Email: kalyan1991n@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

II. NEED FOR DECISION SUPPORT SYSTEM

Several studies have implemented decision support systems in the past to reflect their significance in knowledge management. Some of the prominent ones are enlisted further.

In a study, decision support system is developed to support marketing in business-to-business (B2B) context. The DSS developed helps in management of customer retention in real time scenarios [12]. In another study, RuttOpt a DSS is developed for scheduling of logging trucks in multiple modules. The DSS is adopted by numerous forest and hauling companies resulting in enhanced savings between the ranges of 5-30% [13]. In another study, the importance of integrated DSS (IDSS) is reported from multiple domains focusing on the decision making process. Several cases studies are discussed along with arising challenges [14]. The future of DSS is discussed in another study focusing on development of DSS platforms towards data warehousing, mobile computing and social media systems [15]. In another study, DSS is developed to aid in small city development and management. The system includes tools and techniques for developing novel mathematical tools to help in smart city projects [16]. In another study, bibliographic databases are analyzed to develop clinical DSSs (CDSSs) to aid in medical record data analysis. Furthermore, long term impact of CDSS on genomic data is analyzed [17]. In another study, DSS is designed to support web based applications to generate cluster maps for different applications. The DSS includes tools for analyzing stream of web datasets [18]. T3.3 he performance of spatial DSS (SDSS) is accessed in a study by examining 30 years bibliometric records. Elementary articles and authors are categorized by domain specialization using SDSS which extends the performance of DSS accurately [19].

The significance of DSSs in various domains is highlighted from these studies. Furthermore, designing a twofold DSS for analyzing network based datasets helps in enhancing domain knowledge. The extracted knowledge patterns could be further integrated on multidimensional scale to support real time data interpretation. Hence, the current study focuses on developing a twofold network based DSS.

III. METHODOLOGY

This section highlights the methodology adopted in the study for designing the twofold decision support system.

3.1. Network data collection

a) *Biological network data*: Schizophrenia gene dataset is considered for modelling biological data owing to the prominence of the disease over the years. The gene dataset is curated from multiple schizophrenia gene databases including SZDB [20], SZGR2.0 [21] and DisGeNET [22] based on the biological processes expressed in schizophrenia. b) *Social network data*: Telecommunication sector has experienced wide growth over the years owing to transformation to digital era. Hence, social networking data is collected from Telecom Regulatory Authority of India (TRAI) web portal. Quarterly wireless reports are downloaded from TRAI Web Portal [23].

3.2 Biological network analysis

Schizophrenia gene data collected in previous step is subjected to topic modelling to ascertain relevant gene categories. Supervised Latent Dirichlet Allocation (sLDA) algorithm is employed on the dataset to detect topics. Furthermore, gene communities within the data are detected using novel modularity based approach based on Local Tangent Space Alignment (LTSA) algorithm using 'Rdimtools' package in R programming language [24]. The communities detected are furthermore subjected to tie structure analysis for identifying mutual genes across different biological communities.

3.3 Social network analysis

The wireless dataset downloaded from TRAI portal is subjected to feature selection techniques to identify relevant parameters responsible for successful operation of a telecom player. The features are subjected to Boruta algorithm available in R programming language to detect significant telecom attributes [25]. The features selected from the algorithm are further utilized to detect prominent nodes in the network based on the link based telecom parameters. Prominent telecom players are detected using data mining algorithms on quarterly datasets. The models developed are further validated using statistical metrics like Accuracy and confusion matrix. Based on the outputs, the prominent telecom players are identified.

3.4 Development of network based decision support system

The network based decision support system is developed in R programming language comprising of two interfaces namely: a) *Biological network interface*: The network interface is developed using 'shiny' package available in R programming language [26]. The interface is designed to support data display, topic modelling, community detection and tie strength analysis functionalities.

b) *Social network interface*: Similar to biological interface, the social network interface is developed in 'shiny' dependency. The interface supports telecom data display, feature selection, model development and prominent player's detection.

The DSS designed for biological and social networks is conceptualized by examining the link properties for detecting nodes, hence the name 'BioSocioLink'.

IV. RESULTS AND DISCUSSION

This section highlights the significant outcomes derived from biological and social network analysis.

4.1 Network data processing

The network datasets (i.e. biological and social) are subjected to several analyses prior to development of DSS. The workflow representing these techniques is shown in Fig. 1. The home page of DSS is reflected in Fig. 2.

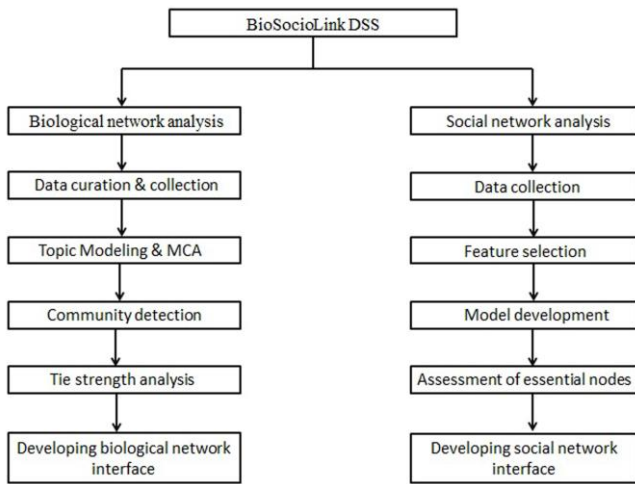


Fig. 1. Workflow of BioSocioLink DSS

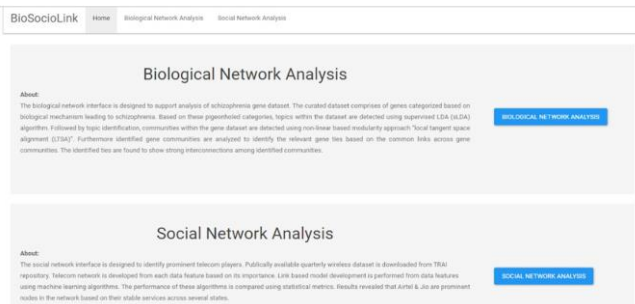


Fig. 2. Home page of BioSocioLink

4.2 Development of biological network interface

Significant knowledge derived from schizophrenia gene data is represented in biological network interface. Initially, the curated dataset is uploaded in the interface followed by accomplishment of topic modeling to detect the strength of gene topics. Out of six gene topics, three topics (i.e. inflammatory, immune response and genetic factors) are found to be highly interrelated compared to others. Furthermore, community detection is implemented by maximizing modularity metric using LTSA algorithm. The algorithm generated six diverse gene communities with enhanced modularity index of 0.92. From the gene communities, interconnected ties are observed throughout the network. These ties are called ‘weak’ owing to less connectivity across different communities. These weak ties act as “hotspots” in identifying relevant schizophrenia causing genes. Each of these functionalities is represented as button events in the DSS. Some of outcomes are represented as Fig. 3, 4, 5 and 6 respectively.

Gene	Category	Function_in_SZ	Published_Reference
CRP	Inflammation	Elevated CRP levels is associated with increased risk of late and very late onset of SZ.	PM64139657
IL2	Inflammation	Inflammatory cytokine associated with activation of cell-mediated immunity in SZ.	7666381
IL6	Inflammation	Inflammatory cytokine associated with activation of cell-mediated immunity in SZ.	7666381
TGFB1	Inflammation	Inflammatory cytokine associated with activation of cell-mediated immunity in SZ.	7666381
IL12	Inflammation	Inflammatory cytokine associated with activation of cell-mediated immunity in SZ.	7666381
IFNG	Inflammation	Inflammatory cytokine associated with activation of cell-mediated immunity in SZ.	7666381
TNF-alpha	Inflammation	Inflammatory cytokine associated with activation of cell-mediated immunity in SZ.	7666381
IL1B	Inflammation	Inflammatory cytokine associated with activation of cell-mediated immunity in SZ.	7666381
COX-2	Inflammation	The proinflammatory protein is associated with pathophysiology in SZ.	15276698
IL8	Inflammation	The inflammatory gene is associated in increased concentrations in serum in SZ.	PM6413205

Fig. 3. Schizophrenia gene dataset displayed with first ten data instances

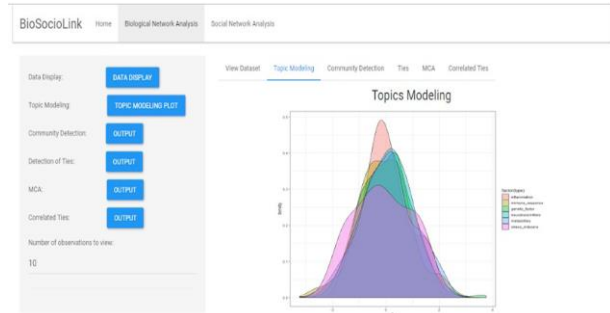


Fig. 4. The output derived from topic modelling representing six prominent gene topics

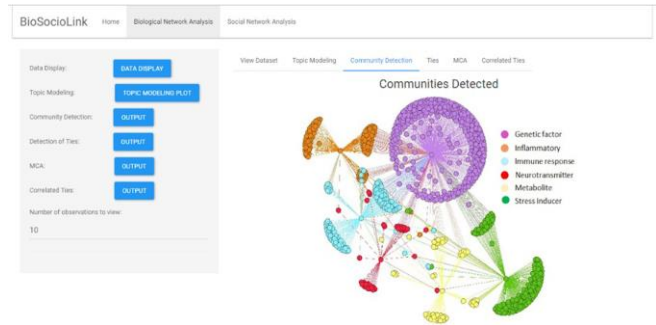


Fig. 5. Schizophrenia gene dataset distributed as six prominent communities

Sl. No	Compressed Gene Ties	Gene Names	Categories they belong to
1	AS1	Protein Inase B	Immune response, Genetic factors, Neurotransmitters, Metabolites
2	DRE2	Dopamine Receptor 2	Genetic factors, Neurotransmitters, Metabolites, Stress inducers
3	IL1B	Interleukin 1B	Inflammatory, Immune response, Genetic factors, Metabolites
4	PPP3CC	Protein Phosphatase 3 Catalytic Subunit Gamma	Immune response, Genetic factors, Neurotransmitters, Metabolites
5	PLA2G4A	Phospholipase A2 Group 4A	Inflammatory, Immune response, Genetic factors, Neurotransmitters
6	SH2B	Syngnase 1-like kinase 1B	Immune response, Genetic factors, Neurotransmitter, Metabolites
7	NO	Nitric oxide	Inflammatory, Immune response, Neurotransmitter, Stress inducer
8	COX2	Cyclooxygenase 2	Inflammatory, Immune response, Genetic factors

Fig. 6. Weak ties interconnecting schizophrenia gene communities

4.3 Development of social network interface

Knowledge patterns derived from telecommunication dataset is represented in social network interface. Initially, 4G wireless quarterly performance datasets are uploaded in the interface as an excel file for each quarter. From the uploaded files, parameters are defined as links and telecom players are defined as nodes. The significance of each link parameter is estimated using feature selection. Boruta algorithm is adopted on every quarterly dataset. Results from these datasets indicated that all the link based parameters are essential in detecting the quality of a telecom operator. Each of these parameters is defined with benchmark indicators from TRAI. Deviation from these indicators will result in imprecise service to consumer. Furthermore, data mining models are developed from each link parameter to access the significant of a telecom operator as per each state. Results revealed that neural network and SVM models outperform other models in identifying prominent players.



Biosociolink: A Decision Support System for Analyzing Link Properties in Biological and Social Networks

Two prominent players (i.e. Airtel & Reliance JIO) are observed in most data patterns. However, emerging players like Vodafone significantly impact the business of these prominent operators due to minimal call drop rates and better service activation at major states. Some of the significant results derived from the analysis is shown in Fig. 7, 8, 9, 10 and 11 respectively.

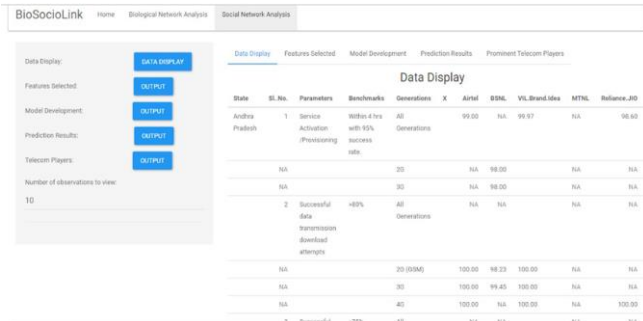


Fig. 7. TRAI Wireless dataset displaying values of telecom operator's w.r.t service activation parameter. Blank cells in the file is displayed as NA

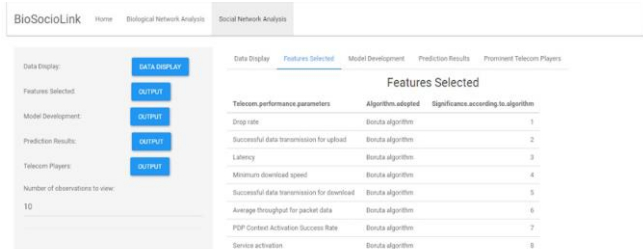


Fig. 8. Feature selection results from Boruta algorithm for March 2019 Quarterly dataset indicating the order of importance of parameters

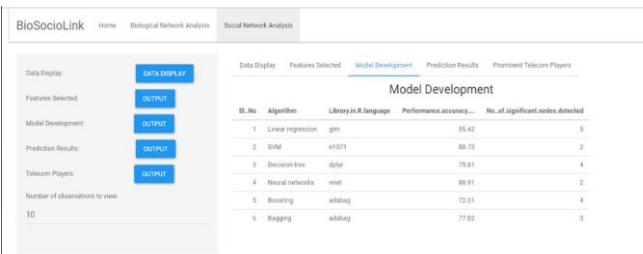


Fig. 9. Results highlighting superior performance of neural net and SVM in detecting prominent telecom operators

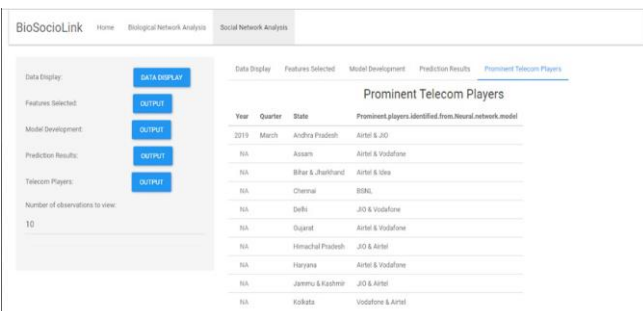


Fig. 10. Significant telecom operators identified by neural network model for March 2019 Quarterly wireless data

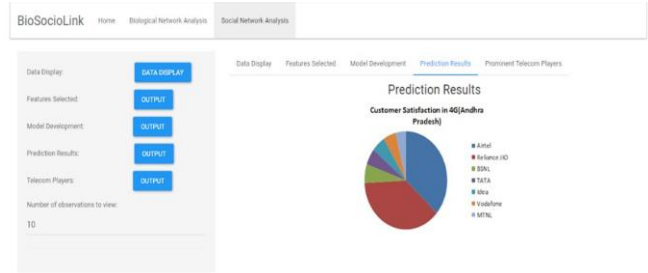


Fig. 11. Pie chart displaying ranking of telecom operators. The chart reflects Airtel & JIO as prominent players for March 2019 dataset in Andhra Pradesh

V. CONCLUSION

The current study focuses on developing a robust decision support system for analyzing network datasets. It aids analyzes of two domains of network data namely biological and social. Primarily, network datasets are imported into the DSS interface and knowledge is extracted. Knowledge is represented in the form of charts and tables. The proposed DSS is deployed in local environment hosted by R programming language. Furthermore, the DSS can be adopted for analysis of human disorders apart from schizophrenia based on common genes expressed among multiple disorders. Alternatively, the GSS can also be adopted on other social networking datasets like friendship, ego, collaboration and citation networks to detect significant connections. However, generation of multi-domain integrative DSS requires advanced analysis. The current study is a preliminary effort in this direction.

ACKNOWLEDGMENT

The authors are grateful to DST-PhD scholarship sponsored by KSTePS, Government of Karnataka and Visveshvaraya Technological University, Jnana Sangama, Belagavi for providing financial assistance to the research work.

REFERENCES

- Watts J, "The "New" Science of Networks" *Annual Review of Sociology*, Volume: 30, 2004, pp. 243-270.
- Nathan Bronson et al, "TAO: Facebook's Distributed Data Store for the Social Graph" *2013 USENIX Annual Technical Conference*, 2013, pp. 49-60.
- Wen Sun, Achille Fokoue, Kavitha Srinivas, Anastasios Kementsietsidis, Gang Hu, Guotong Xie, "SQLGraph: An Efficient Relational-Based Property Graph Store" *ACM SIGMOD*, 2015, pp. 1-15.
- Prithviraj Sen, Galileo Namata, Mustafa Bilgic, Lise Getoor, Brian Galligher, Tina Eliassi-Rad, "Collective classification in network data" *AI Magazine*, Volume: 29, Issue: 3, 2008, pp. 1-24.
- Linyuan Lu, Tao Zhou, "Link prediction in complex networks: A survey" *Physica A: Statistical Mechanics and its Applications*, Volume: 390, Issue: 6, 2011, pp. 1150-1170.
- Santo Fortunato, "Community detection in graphs" *Physics Reports*, Volume: 486, Issue: 3-5, 2010, pp. 75-174.
- Eric Alm, Adam P Arkin, "Biological networks" *Current Opinion in Structural Biology*, Volume: 13, Issue: 2, 2003, pp. 193-202.
- Xiaowei Zhu, Mark Gerstein, Michael Snyder, "Getting connected: analysis and principles of biological networks" *Genes & Dev.*, Volume: 21, 2007, pp. 1010-1024.
- Degenne A, Forse M, "Introducing Social Networks" *SAGE Publications*, 1999.

10. Jennifer Golbeck, "The dynamics of Web-based social networks: Memberships, relationships and change" *First Monday*, Volume: 12, Issue: 11, 2007.
11. Daniel Power, "Decision Support Systems: From the Past to the Future" *Americas Conference on Information Systems (AMCIS)*, 2004, pp. 2025-2031.
12. Noori B, Hossein Salimi M, "A decision-support system for business-to-business marketing" *Journal of Business & Industrial Marketing*, Volume: 20, Issue: 4-5, 2005, pp. 226-236.
13. Andersson Gert, Flisberg Patrik, Lidén Bertil, Rönqvist Mikael, "RuttOpt – A decision support system for routing of logging trucks" *NHH Dept. of Finance & Management Science Discussion Paper*, 2007, pp. 1-35.
14. Shaofeng Liu, Alex H. B. Duffy, Robert Ian Whitfield, Iain M Boyle, "Integration of decision support systems to improve decision support performance" *Knowledge and Information Systems*, Volume: 22, Issue: 3, 2010, pp. 261-286.
15. Bryan Hosack, Dianne Hall, David Paradice, James F. Courtney, "A Look Toward the Future: Decision Support Systems Research is Alive and Well" *Journal of the Association for Information Systems*, Volume: 13, 2012, pp. 315-340.
16. Virgil Chichernea, "The Use Of Decision Support Systems (Dss) In Smart City Planning And Management" *Romanian Economic Business Review*, Volume: 8, Issue: 2, 2014, pp. 238-251.
17. J Bouaud, V Koutkias, "Computerized Clinical Decision Support: Contributions from 2014" *Yearbook of Medical Informatics*, Volume: 10, Issue: 1, 2015, pp. 119-124.
18. Peter Kennan, "Changes in DSS disciplines in the Web of Science" *Journal of Decision System*, Volume: 26, Issue: 1, 2016, pp. 542-549.
19. Peter Bernard Keenan, Piotr Jankowski, "Spatial Decision Support Systems: Three decades on" *Decision Support Systems*, Volume: 116, 2019, pp. 64-76.
20. Wu Y, Yao Y.-G, Luo X.-J, "SZDB: A Database for Schizophrenia Genetic Research" *Schizophrenia Bulletin*, Volume: 43, 2017, pp. 459-471.
21. Jia P, Han G, Zhao J, Lu P, Zhao Z, "SZGR 2.0: a one-stop shop of schizophrenia candidate genes" *Nucleic Acids Research*, Volume: 45, 2017, pp. D915-D924.
22. Piñero J, Bravo À, Queralt-Rosinach N, Gutiérrez-Sacristán A, Deu-Pons J, Centeno E, García-García J, Sanz F, Furlong LI, "DisGeNET: a comprehensive platform integrating information on human disease-associated genes and variants" *Nucleic Acids Research*, Volume: 45, 2017, pp. D833-D839.
23. <https://main.traf.gov.in/release-publication/reports/wireless-data-reports>
24. <https://cran.r-project.org/web/packages/Rdimtools/Rdimtools.pdf>
25. <https://cran.r-project.org/web/packages/Boruta/Boruta.pdf>
26. <https://cran.r-project.org/web/packages/shiny/shiny.pdf>

AH Manjunatha Reddy is currently working as Associate Professor in department of Biotechnology at RV College of Engineering. He has completed his PhD in Environmental Science under Kuvempu University. Prior to this he has completed Master of Science and Bachelor of Science under Kuvempu University. His research interest lies in bioinformatics, environmental biotechnology and biostatistics. He has completed several funded consultancy projects under his supervision.



Kalyan Nagaraj is currently pursuing PhD in Computer Science & Engineering at RV College of Engineering. Prior to this she has completed M.Tech & BE from Visveshvaraya Technological University, Belgaum. His research interests broadly include data science, network security and computational biology.

AUTHORS PROFILE



Amulyashree Sridhar is currently pursuing PhD in Computer Science & Engineering at RV College of Engineering. Prior to this she has completed M.Tech & BE from Visveshvaraya Technological University, Belgaum. Her research interests are network science, data science and bioinformatics.



Sharvani GS is currently working as an associate professor in Department of Computer Science, RV College of Engineering. She has completed her PhD in Computer Science from Avinashi Lingam University For Women, Coimbatore. She has completed her Masters in Technology in Computer Networks and B.E from Visveshvaraya Technological University. Her research interest lies in cloud computing, internet of things (IoT), swarm intelligence, grid computing. She has supervised several consultancy and funded projects till date.



Ramakanth Kumar P is working as Professor and Head of Computer Science and Engineering department at RV College of Engineering. He has completed his PhD in Computer Science and Engineering from Mangalore University. He has vast experience of 25 years in teaching and research. He has been the chairperson of several national and international conferences. He is nominated for Board of Studies in several reputed institutes across the state. He has supervised several PhDs under his guidance. His research interests broadly falls under pattern recognition, natural language processing and digital image processing.



Reference: A1914109119/2019©BEIESP
DOI: 10.35940/ijeat.A1914.109119
Journal Website: www.ijeat.org