

# **2016 International Conference on Data Science and Engineering (ICDSE 2016)**

**Cochin, India  
23 – 25 August 2016**



**IEEE Catalog Number: CFP1627T-POD  
ISBN: 978-1-5090-1282-4**

**Copyright © 2016 by the Institute of Electrical and Electronics Engineers, Inc  
All Rights Reserved**

*Copyright and Reprint Permissions:* Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

***\*\*\*This publication is a representation of what appears in the IEEE Digital Libraries. Some format issues inherent in the e-media version may also appear in this print version.***

IEEE Catalog Number:	CFP1627T-POD
ISBN (Print-On-Demand):	978-1-5090-1282-4
ISBN (Online):	978-1-5090-1281-7

**Additional Copies of This Publication Are Available From:**

Curran Associates, Inc  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: (845) 758-0400  
Fax: (845) 758-2633  
E-mail: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

CURRAN ASSOCIATES INC.  
**proceedings**  
.com

# Table of Contents

ICDSE 2016

## 2016 International Conference on Data Science & Engineering

### Welcome Message

### Conference Organization

### Session 1: Data Processing and Big data Infrastructure

Improvement of ETL through Integration of Query Cache and Scripting method <i>Prayag Tiwari</i>	3
Data Preparation: Art or Science? <i>Gunjan Mansingh, Kweku-Muata Osei-Bryson, Lila Rao and Maurice McNaughton</i>	8
Generalized Approach to Linear Data Transformation <i>Abhijith Asok</i>	14
Not Just Dissimilar, but Opposite, An Algorithm for Measuring Similarity and Oppositeness between Words <i>Dean J. Jones and Gunjan Mansingh</i>	20
Understanding the Indian Labour Market: A Data-Centric Approach <i>Shabana K M, Tony Gracious and Hrishikesh Subramonian</i>	26
MapReduce Model For K-Medoid Clustering <i>Sandhya Harikumar, Shabana Shamsuddin Thaha</i>	32

### Session 2: Cloud Computing and Data Management

Auto Scaling for Various Patterns of Workflow within Deadline time and Energy aware VM allocation in Cloud Environment <i>Sruthy Santhosh and Binu A.</i>	41
--	----

SADA : Secure Approximate Data Aggregation in Wireless Sensor Networks	46
<i>G Prathima E, Shiva Prakash T, Venugopal K R, S S Iyengar and L M Patnaik</i>	
<i>Erasure Coded Storage Systems for Cloud Storage Challenges and Opportunities.</i>	52
<i>Ojus Thomas Lee, S.D Madhu Kumar and Priya Chandran</i>	
Clustering of Users on Microblogging Social Media: A Rough Set Based Approach	59
<i>Mukul Gupta, Pradeep Kumar and Bharat Bhasker</i>	
 <b>Session 3: Machine Learning Techniques and Systems</b>	
An Abstract Framework for Ontology Evaluation	67
<i>Amandeep Verma</i>	
Comparative study of Data Mining Clustering algorithms.	72
<i>Iyer Aurobind Venkatkumar and Sanatkumar Jayantibhai Kondhol Shardaben</i>	
Student Profiling to improve teaching and learning: A DataMining Approach	79
<i>Anand Desai, Nemil Shah and Madhuri Dhodi</i>	
Breast Cancer Detection using Two-Fold Genetic Evolution of Neural Network Ensembles	N/A
<i>Indu Singh, Karan Sanwal and Satyarth Praveen</i>	
Ranking of Suppliers with MCDA Technique and Probabilistic Criteria	91
<i>Susmita Bandyopadhyay</i>	
Teacher Recruitment Data Analytics using Association Rule Mining in Indian Context	96
<i>Vijayakumar Kadappa, Shankru Guggari and Atul Negi</i>	

Machine Learning Algorithms for Document Clustering and Fraud Detection	103
<i>Suresh Yaram</i>	
Tree-Based Ensemble Models for Predicting the ICU Transfer of Stroke In Patients	109
<i>Nawal N. Alotaibi and Sreela Sasi</i>	
Apriori Algorithm for Association Rule Mining in High Dimensional Data	115
<i>Sandhya Harikuma and Divya Usha Dilipkumar</i>	
A Statistical Analysis of Soil Fertility of Thrissur District, Kerala	121
<i>R. Ajith kumar, M.K. Muhammed Aslam, Jagathy Raj V.P, T. Radhakrishnan, K. Satheesh kumar, and T. K. Manojkumar</i>	
PTMIB: Profiling Top Most Influential Blogger using Content Based Data Mining Approach	126
<i>Vasanthakumar G U, Priyanka R, Vanitha Raj K C, Bhavani S, Asha Rani B R, P Deepa Shenoy and Venugopal K R</i>	

#### **Session 4: Information Storage, Search and Retrieval**

Study on Wavelet Spectral Band Based EEG Compression	135
<i>C R Midhun Raj, Harsha A</i>	
Dynamic Prediction Based Watermarking for Temporal Relational Databases	140
<i>Unnikrishnan K and Pramod K V</i>	
Proper Imputation Techniques for Missing Values in Data sets	146
<i>Tahani Aljuaid and Sreela Sasi</i>	
A Local Search Approximation Algorithm for the Restricted Universal Facility Location Problem	151
<i>Uttam Kanodia and K K Shukla</i>	

#### **Session 5: Social Networks and Personal Information**

Anomaly Detection in Web Graphs using Vertex Neighbourhood based Signature Similarity methods	157
<i>Aritra Ghosh and Pallavi Gudipati</i>	

Inferring Borrower Network in a Microfinancing Framework (KIVA)	<b>163</b>
<i>Aritra Ghosh and Jithin Vachery</i>	
Detection of Similar Community in Large Network Based on Graph Edit Distance	<b>168</b>
<i>Abhay Narayan and G. Santhosh Kumar</i>	
An Enhanced Model for Behavioral Targeting in Online Advertising	<b>173</b>
<i>Radhika V M, Aiswarya Thottungal and Abdul Nizar M</i>	
Using mobile phone data to study dynamics of rural-urban mobility	<b>178</b>
<i>Rahman Sanya and Martin Mubangizi</i>	
Community Detection in Heterogenous Networks Using Incremental Seed Expansion	<b>184</b>
<i>Samrat Gupta and Pradeep Kumar</i>	
<b>Session 6: Computer Networks, Privacy and Security</b>	
User Characterization through Network Flow Analysis	<b>191</b>
<i>Vinupaul M.V, Raktim Bhattacharjee, Rajesh R and G. Santhosh Kumar</i>	
Search Control in Multi-user Searchable Encryption without Shared keys	<b>197</b>
<i>Manju S Nair and Rajasree M S</i>	
Cooperative Relaying for Interference Mitigation in Device to Device communication	<b>202</b>
<i>Minu P N and Premkumar C V</i>	
SWCTE:Semantic Weighted Context Tagging Engine for Privacy Preserving Data Mining	<b>207</b>
<i>N P Nethravathi, Prasanth G Rao, Vaibhav J Desai, P Deepa Shenoy, Venugopal K R and Indiramma M</i>	

## **Author Index**