

Third Workshop on Data Mining Case Studies and Practice Prize

Call For Papers

Third Workshop on Data Mining Case Studies and Success Stories
and Third Data Mining Practice Prize

Motivation

From its inception the field of data mining has been guided by the need to solve practical problems. Yet a cursory examination of the publications shows that few papers describe a completed implementation or what we will term a “case study”. The small number of case studies is counter-balanced by their prominence. Anecdotally case studies are one of the most discussed topics at data mining conferences. Some of the benefits of good case studies include

1. *Inspiration: Case studies provide examples that can inspire data mining researchers to pursue important new technical directions.*
2. *Innovation: Data mining case studies demonstrate how whole problems were solved - not just part of the problem. Often building the prediction algorithm is only 10% of the problem - the other aspects that comprise a successful deployment are valuable for practitioners to understand.*
3. *Education: People are more likely to remember stories than facts.*
4. *Media Coverage: The media is more likely to report on completed data mining applications, than they are on isolated algorithms. We have an opportunity to present positive success stories to the wider community.*
5. *Public relations: Applications, particularly those that are socially beneficial, will help our perception both within the wider public and other scientific fields.*
6. *Connections to Other Scientific Fields: Completed systems knit together a range of scientific and engineering disciplines such as signal processing, chemistry, optimization theory, auction theory and so on. Fostering meaningful connections to these fields will benefit data mining academically, and will assist data mining practitioners to learn how to harness these fields to develop successful applications.*

The Workshop

The Data Mining Case Studies Workshop and Practice Prize was established to showcase the very best in data mining case deployments. Data Mining Case Studies continues into its third year, to be held at [KDD2009](#). Data Mining Case Studies will highlight data mining implementations that have been responsible for a significant and measurable

improvement in business operations, or an equally important scientific discovery, or some other benefit to humanity.

Examples of Data Mining Case Studies from previous years have included: (a) a medical application that has saved hundreds of lives by mining through hundreds of thousands of patient records to identify patients who have shown all the signs for heart disease, yet have not been prescribed heart medication, (b) a system which has uncovered hundreds of millions in sheltered tax evasion rings, (c) a system which has raised revenue by improved cross-selling of computer peripherals and equipment.

Data Mining Case Studies will allow papers greater latitude in (a) range of topics - authors may touch upon areas such as optimization, operations research, inventory control, and so on, (b) page length - longer submissions are allowed, (c) scope - more complete context, problem and solution descriptions will be encouraged, (d) prior publication - if the paper was published in part elsewhere, it may still be considered if the new article is substantially more detailed, (e) novelty – the use of established techniques to achieve successful implementations will be given partial allowance.

Unsuccessful data mining systems that describe lessons learned and “war stories” will also be assessed.

The Data Mining Practice Prize

Introduction: The Data Mining Practice Prize will be awarded for the best Data Mining Case Study submission. The prize will be awarded for work that has had a significant and quantitative impact in the application in which it was applied, or has significantly benefited humanity. Detailed rules and regulations will be finalized upon workshop acceptance.

Eligibility: All papers submitted to Data Mining Case Studies will be eligible for the Data Mining Practice Prize, with the exception of the Data Mining Practice Prize Committee. Eligible authors must consent to allowing the Practice Prize Committee independently validate their claims by contacting third parties and their deployment client for independent verification and analysis.

Award: Winners and runners up can expect an impressive array of honors including

1. Article summaries about each of the deployments to be published in the journal [SIGKDD Explorations](#), which will also announce the results of the competition and the prize winners.
2. Prize money comprising \$500 for first place, \$300 for second place, \$200 for third place.
3. Plaque.
4. Awards Dinner with organizers and prize winners.

Topics

Most operational industrial and scientific systems that involve data mining to some extent are likely to be acceptable. Systems that are responsible for mission critical systems, medical applications, cash flow, or applications that significantly benefit humanity will be particularly good candidates. If you are unsure as to the suitability of your paper, please contact the organizers with your topic at the email address at the bottom of the page. Topics include but are not limited to

- *Genomics*
- *Inventory control*
- *Customer Relationship Management (CRM)*
- *ShopBots*
- *Recommendation systems*
- *Auction trading systems*
- *Clinical patient monitoring*
- *Seismic Data interpretation*
- *Survival analysis for medical procedures*
- *Climate analysis*
- *Correlates of genes with disease*
- *Dangerous Drug interactions*
- *Law enforcement applications*
- *Search Engine Marketing*
- *Food spoilage elimination*
- *Price optimization*
- *Data visualization in mission-critical user interfaces*
- *Text understanding*

Dates

Submissions open	Feb 21, 2009
Notify organizers of intent to submit	now
Optional Draft submission including client contact information. The draft is requested for planning purposes*	March 30, 2009
Final submission including client contact information if it has not already been provided	April 8, 2009
Notification of acceptance	May 15, 2009
Camera ready paper submission	May 25, 2009
Workshop delivers papers to IEEE for printing	May 29, 2009
Workshop held, Practice Prize winners announced	June 28, 2009

We recommend authors submit a draft of their paper by March 30 2009 so that we can better begin the process of validating claims. Only the chairs will see the paper - it will not be seen by the reviewers. If you cannot provide the article by this date, you may still submit by the final submission deadline.

Submission instructions

In order to contact the organizers, submit, or for any other correspondence, please use the following email address

(note that the email address which follows has been encoded as a bitmap)

submissions@dataminingcasestudies.com

1. Please email the organizers as early as possible with your intention to submit.
2. If possible, it is recommended that you provide an optional draft of the article by the draft submission date. This draft will only be viewed by the Chairs - it will not be given to the reviewers or affect the prize competition.
3. Please provide us with three persons who use the system in their day to day activities, or are responsible for the system, and who may be contact to validate the claims made in the paper. Ideally these individuals belong to a different company than the authors. Also, ideally these individuals are not personal acquaintances or friends of the authors.
4. Provide your author names, addresses, affiliations, phone numbers and email. Also note the nature of relationship of each contact to the system and authors. Finally, provide any information of relevance to contacting deployment users.
5. Please submit your completed article, in ACM SIGKDD Proceedings format to the email address above. Due to editing requirements for the Workshop Proceedings, we strongly encourage documents to be submitted in Microsoft Word format. The official [ACM Microsoft Word template](#) is available from [here](#).

Guidelines

1. Word limits: Word limits will be relaxed for submission to Data Mining Case Studies, so that participants may explain their problem and solution in as much detail as necessary to both captivate the reader and explain the solution. The maximum submission page length will be 20 pages. Despite the longer page length, articles will be critically assessed for relevancy, and authors risk rejection if their articles do not keep the reader's interest. In addition, the PC will look for ways to cut the article, and so any recommendations made by the PC for cutting the article will need to be followed to prior to inclusion in the workshop program
2. Commercial product mentions: Data Mining Case Studies is not a sales venue. References to commercial products will be carefully scrutinized by our Program Committee for applicability. Where possible the underlying techniques should be described. The purpose of Data Mining Case Studies is to illustrate real applications with descriptions that are concise and complete. Commercial software if introduced, should be named briefly and then described at a technical

- level (eg. don't mention that "SAS Neural Nets(TM) increased our forecast accuracy by 20%" - instead say that you used 'SAS PROC Neural Net(TM)' which implemented a 3- layer sigmoidal backpropagation model with 10 inputs, 4 hidden and 1 output node, and this net increased forecast accuracy by 20% ". Any papers violating these ethics will be deemed inadmissible. If in doubt please contact the organizers prior to submission. We will allow a single product mention along the lines described above, and this should be sufficient for establishing commercial credibility.
3. Valid contact information for the company that deployed the data mining system must be supplied to the Program Committee. The Program Committee should be afforded the right to contact individuals that were the beneficiaries of the data mining system and ask them questions about the implementation. In particular, the claims made in the paper submission will need to be verified. Failure to provide factual or complete descriptions of results obtained with the system, that are discovered through this fact checking process, will result in forfeiture of prize and dismissal from the conference. The Prize Committee will endeavor to be discrete in its contacts, so please inform us of any information we need to know before contacting the system users.
 4. Copyright: Authors will agree to allow the display of their articles on the web. Authors should also agree to allow their articles to be published in book form. If authors wish to opt out of website or book publication, please contact the Workshop organizers.
 5. Confidentiality: The reviewing process will be confidential.

Venue

15th International Conference on Knowledge Discovery and Data Mining, Paris, France, June 28-July 1, 2009

Honors for Accepted Papers

1. Accepted papers will be published in Data Mining Case Studies Workshop by ACM on CDROM.
2. Accepted authors will be invited to give a presentation at 15th International Conference on Knowledge Discovery and Data Mining in Paris, France, June 28-July 1, 2009
3. Prize winners will write a short overview of their article for the journal SIGKDD Explorations
4. Prize winners receive a plaque.
5. Successful papers may be published in a "Data mining success stories" book

Organizing Committee

Gabor Melli (co-chair), Predictionworks
Peter van der Puttem (European co-chair), PhD., Chordiant Software
Brendan Kitts (co-chair), Microsoft
Gregory Piatetsky-Shapiro, PhD., KDNuggets
Karl Rexer, PhD., Rexer Analytics
Gang Wu, PhD. Microsoft
Dean Abbott, PhD., Abbott Analytics
Venky Ganti, PhD., Microsoft
Jing Ying Zhang, PhD. Microsoft
Arber Xu, PhD., Summba
David Duling, PhD., SAS Corporation
Kunal Punera, PhD., Yahoo! Inc.
Moninder Singh, PhD., IBM Corporation
Felix Chen, Predictionworks,
Richard Bolton, PhD., KBMI
Robert Grossman, PhD., University of Illinois at Chicago
Ed Freeman, DS-IQ

Further Information

<http://www.dataminingcasestudies.com>

Gold Sponsors

