ISO/IEC *AI workshop sessions* on 29 and 30 November

ISO/IEC 24668
Process Management Framework - Big Data Analytics

GAUTAM BANERJEE
BUSINESS BRIO

Session 2
30th November 2022
The Data Science Struggle: Why Organizations Routinely Fail To Realize The Full Potential From Their Data Science Efforts

Hannah M. Mayer Contributor

By Hannah M. Mayer, Luca DeStefano

Nov 7, 2022, 03:30am EST
Data-driven decision making will fail – and here is why

Marc Warner, CEO at Faculty, contends that corporate organisations need to go through a cycle similar to the development of scientific thought – from divination, via data deluge to theoretical understanding.

Most people have heard of data-driven decision making – the admirable desire to make better decisions through data. The conventional justification is that given the world’s complexity, organisations need data to make decisions.

Such ideas exist across organisations from governments to companies to hospitals. In fact, they are wrong in principle and misleading in practice.

Data-driven decision making puts the emphasis in the wrong place, and means organisations focus on the wrong thing. Let’s look at why.
How to Prevent Big Data Analytics Failures

December 18, 2015

Contributor: Susan Moore

Big data analytics projects don't fail for a single reason, nor due to technology alone.
“To succeed, you must develop a viable strategy to deliver business value from a big data initiative. Then map out and acquire or develop the missing and specialized skills that are needed. Once strategy and skill priorities are addressed, then you can move on to big data analytics.”

Gartner research director Svetlana Sicular

Standards

Need
Sharing; Re-use; Consent; Collaborate; Interoperability; Adoption; Trust

Forms
Technology; Frameworks; Methodology; Use Cases

Means
Published papers; Patents; Standards; Certification; Regulation; Legislation

Source
Forums; Bodies; National; International; Unions and groups of countries
4 Standards
From ISO/IEC SC 42

B  5259 Series
   Data Quality

C  5338
   AI LifeCycle Methodology

C  38507
   Governance implications for AI applications

A  24468
   Process Management Framework - Big Data Analytics

NWIP | WD | CD | DIS | FDIS | IS
Process Management Framework - Big Data Analytics

The proposed standard provides a framework for developing processes to effectively leverage big data analytics across the organization irrespective of the industries/sectors. It specifies process management for big data analytics with the five process categories as listed below with help of Process Reference Model, Measurement Framework and Assessment Model:

- Context
- Purpose
- Outcomes
- Attributes
- Rating Scale
- Process Capability by processes from Process groups
**Sample process Reference Model**

A sample of Data Identification process (refer DMP1 from previous figure) from Data Management Process group is given below for understanding.

<table>
<thead>
<tr>
<th>ID</th>
<th>DMP1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Data identification</td>
</tr>
<tr>
<td>Context</td>
<td>This process covers the key step of identifying data elements. This involves identification of data elements so that the team will not lose sight of any important dimension or factor that plays a key role in the outcome analysis.</td>
</tr>
<tr>
<td>Purpose</td>
<td>The purpose of DMP1 process is to identify, define, classify, and collect data for all data elements available for the information flow in the context of the project or department or function.</td>
</tr>
</tbody>
</table>
| Outcomes | The outcomes of this process include:  
  a) data elements relevant to the process or function or department are identified;  
  b) data elements can be classified into categories, such as unstructured, transactional, hierarchical, reference;  
  c) meta data (operational definition) such as units, frequency, source or sources, functional definition, range, possible functional correlation, producer or consumer or ownership or steward is derived;  
  d) new data sources are identified;  
  e) data collection is performed. |
Sample process Base Practices and Information Products

A sample structure for defined base practices for Data Identification process is given below

<table>
<thead>
<tr>
<th>ID</th>
<th>DMP1</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Base practices**

BP1 identify the key data elements: such as features, co-variates, factors [outcome (a,d)].
BP2 classify the data elements: [outcome (b)].
BP3 provide the operational definition of data elements: [outcome (c, d)].
BP4 collect the data with rigor and compliance: [outcome (e)].

A sample structure of Information Products (IPs) for Data Identification process is given below

<table>
<thead>
<tr>
<th>IP ID</th>
<th>Name</th>
<th>Outcome</th>
<th>IP ID</th>
<th>Name</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP_27</td>
<td>SIPOC</td>
<td>a), b)</td>
<td>IP_32</td>
<td>Data collection plan</td>
<td>a), c), d), e)</td>
</tr>
<tr>
<td>IP_28</td>
<td>Process flow</td>
<td>b), c)</td>
<td>IP_33</td>
<td>Data operational definition</td>
<td>a), b), c), d), e)</td>
</tr>
<tr>
<td>IP_29</td>
<td>CTQ matrix</td>
<td>a), b), c)</td>
<td>IP_34</td>
<td>Data sample</td>
<td>e)</td>
</tr>
<tr>
<td>IP_30</td>
<td>Use case studies</td>
<td>c), d)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IP_31</td>
<td>Domain documents</td>
<td>c), d)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Assessment

For each process, the process assessment model further provides:

- A set of base practices (BPs) defining the tasks and activities needed to accomplish the process purpose and fulfil the process outcomes; each base practice is explicitly associated to a process outcome.
- Number of information products (IPs) associated with each process and related to one or more of its outcomes.
- Characteristics associated with each information products.

The base practices and information products constitute the set of indicators to assess process performance for a particular process. The capability levels of each process area can be determined accordingly and then action items can be drafted to address the gaps.

More importantly, the assessment should lead to a roadmap to implement Data Science/AI or Big Data Analytics in terms of use cases for different functions.
Thank You

Gautam Banerjee

https://www.linkedin.com/in/gautambanerjee/
Appendix
5259 Series

Data Quality

MEASURES
• Accuracy
• Precision
• Uniqueness
• Completeness
• Consistency

CHALLENGES
• Structured Data vs BigData
• ML vs AI
• Part 4 vs Data Governance

* These tenets are WIP. They may not be accurate and meant to give only a high level idea
AI Lifecycle Methodology

Start → Requirements analysis → Architecture definition design → System analysis
Integration → Implementation → Data engineering → Knowledge Acquisition
Verification → Transition → Validation → Continuous monitoring model updates

* These tenets are WIP. They may not be accurate and meant to give only a high level idea.
Governance implications for AI applications

* These tenets are WIP. They may not be accurate and meant to give only a high level idea.