Deep Learning
Indaba X - Zambia 2021

Using Machine Learning Techniques for Solving Locally Relevant Problems

Lighton Phiri <lighton.phiri@unza.zm>
Department of Library & Information Science
University of Zambia
http://lis.unza.zm/~lightonphiri
About The DataLab Research Group at The University of Zambia

- The DataLab research group at The University of Zambia is composed of faculty staff and students—undergraduate and postgraduate—working in three main areas
  - Data Mining
  - Digital Libraries
  - Technology-Enhanced Learning

Data Mining
With the proliferation of data, the field of Data Mining has gained rapid popularity. Data Mining focuses on the discovery of patterns in large datasets by making use of statistical and machine learning techniques.

Our current focus involves leveraging machine learning techniques to facilitate efficient and effective delivery of services in the health and educational domains—two areas that are of significance in the so-called developing world.

The following members are actively working in this area:

Digital Libraries
The field of Digital Libraries (DLs) generally involves the study of digital collections of information and corresponding network-based services used to retrieve data from the collections. DLs are in effect information systems that are used to persistently store digital objects, manage the digital objects and, facilitate access to digital objects.

Our focus in the field of DLs, as a research group, mostly involves experimenting with techniques that can potentially facilitate efficient and effective access to digital objects stored in DLs.

The following members are actively working in this area:
Outline

- Part I. Data-Driven Problem Solving
- Part II. Past and Current Projects
- Part III. Potential Problems
Outline

● Part I. Data-Driven Problem Solving
  ○ Introduction
  ○ Data Mining Pipelines
  ○ Data Mining Models

● Part II. Past and Current Projects

● Part III. Potential Problems
Machine Learning 101 [...]

- Artificial Intelligence encompasses a broad spectrum of sub-fields
  - Traditional machine learning techniques and approaches
  - Deep Learning approaches
Machine Learning 101 [...] 

- Artificial Intelligence encompasses a broad spectrum of sub-fields
  - Traditional machine learning techniques and approaches
  - Deep Learning approaches
Data is Key to ML-Centric Problem Solving
Data Mining Pipelines

- Fundamentally, machine learning aims to extract knowledge from data
  - Historical data is used to infer/predict outcomes associated with new observations
Data Mining Pipelines

- Input features identified during feature engineering are used to train models
  - Features correlated with outcome to be identified
Data Mining Pipelines

- The ML inference model is used to predict future patterns
  - Models can then be deployed as Web services and/or standalone applications
Data Mining Models (1/5)

- Numerous data mining models and frameworks have been proposed
  - Most trace their roots from the KDD Process proposed by Fayyad et al.

https://doi.org/10.1017/S0269888910000032
Data Mining Models (2/5)
Data Mining Models (2/5)

[Diagram showing the process of data mining, from data selection, preprocessing, transformation, mining, to interpretation and evaluation, leading to knowledge.]
## Data Mining Models (2/5)

<table>
<thead>
<tr>
<th>Methodology</th>
<th>CRISP-DM</th>
<th>Business understanding</th>
<th>Data understanding</th>
<th>Data preparation</th>
<th>Modeling</th>
<th>Evaluation</th>
<th>Deployment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human-Centered</td>
<td>Task discovery</td>
<td>Task discovery</td>
<td>Data cleaning</td>
<td>Model development</td>
<td>Data analysis</td>
<td>Data analysis</td>
<td>Output generation</td>
</tr>
<tr>
<td>Human-Centered</td>
<td>Data discovery</td>
<td>Data discovery</td>
<td>Model development</td>
<td>Data analysis</td>
<td>Assess</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methodology</th>
<th>SEMMA</th>
<th>Sample</th>
<th>Explore</th>
<th>Modify</th>
<th>Assess</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cabena et al.</td>
<td>Select</td>
<td>Preprocess</td>
<td>Pre-process</td>
<td>Transform</td>
<td>Mine</td>
</tr>
<tr>
<td>Cabena et al.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Analyse &amp; assimilate</td>
</tr>
<tr>
<td>Two Crows</td>
<td>Define business problem</td>
<td>Build DM data base</td>
<td>Explore data for modeling</td>
<td>Build model</td>
<td>Evaluate model</td>
</tr>
<tr>
<td>Two Crows</td>
<td>Explore data</td>
<td>Prepare data</td>
<td>Build model</td>
<td></td>
<td>Deploy model and results</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Methodology</th>
<th>Anand &amp; Buchner</th>
<th>Domain knowledge elicitation</th>
<th>Domain knowledge elicitation</th>
<th>Methodology identification</th>
<th>Methodology identification</th>
<th>Knowledge post-processing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anand &amp; Buchner</td>
<td>Human resource identification</td>
<td>Data prospecting</td>
<td>Data pre-processing</td>
<td>Pattern discovery</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anand &amp; Buchner</td>
<td>Problem specification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DOI: https://doi.org/10.1017/S0269888910000032
Data Mining Models (3/5)

Relative effort (%)

- Cabena et al. estimates
- Shearer estimates
- Cios and Kurgan estimates

KDDM steps

https://doi.org/10.1017/S0269888906000737
Data Mining Models (4/5)

- CRISP-DM model is one of the most widely used data mining models
- Data understanding and preparation are the most time consuming

What main methodology are you using for your analytics, data mining, or data science

<table>
<thead>
<tr>
<th>Methodology</th>
<th>2014 poll</th>
<th>2007 poll</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRISP-DM (86)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>My own (55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SEMMA (17)</td>
<td>8.5%</td>
<td>13%</td>
</tr>
<tr>
<td>Other, not domain-specific (16)</td>
<td>8%</td>
<td>4%</td>
</tr>
<tr>
<td>KDD Process (15)</td>
<td>7.5%</td>
<td>7.3%</td>
</tr>
<tr>
<td>My organizations' (7)</td>
<td>3.5%</td>
<td>5.3%</td>
</tr>
<tr>
<td>A domain-specific methodology (4)</td>
<td>2%</td>
<td>4.7%</td>
</tr>
<tr>
<td>None (0)</td>
<td>0%</td>
<td></td>
</tr>
</tbody>
</table>

https://www.kdnuggets.com
Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology

Stefan Studer, Thanh Binh Bui, Christian Drescher, Alexander Hanuschkin, Ludwig Winkler, Steven Peters, Klaus-Robert Mueller

Machine learning is an established and frequently used technique in industry and academia but a standard process model to improve success and efficiency of machine learning applications is still missing. Project organizations and machine learning practitioners have a need for guidance throughout the life cycle of a machine learning application to meet business expectations. We therefore propose a process model for the development of machine learning applications, that covers six phases from defining the scope to maintaining the deployed machine learning application. The first phase combines business and data understanding as data availability oftentimes affects the feasibility of the project. The sixth phase covers state-of-the-art approaches for monitoring and maintenance of a machine learning applications, as the risk of model degradation in a changing environment is eminent. With each task of the process, we propose quality assurance methodology that is suitable to address challenges in machine learning development that we identify in form of risks. The methodology is drawn from practical experience and scientific literature and has proven to be general and stable. The process model expands on CRISP-DM, a data mining process model that enjoys strong industry support but lacks to address machine learning specific tasks. Our work proposes an industry and application neutral process model tailored for machine learning applications with focus on technical tasks for quality assurance.


Subjects: Machine Learning (cs.LG); Software Engineering (cs.SE); Machine Learning (stat.ML)

Cite as: arXiv:2003.05155 [cs.LG]
(or arXiv:2003.05155v2 [cs.LG] for this version)

Submission history
From: Thanh Binh Bui [view email]
Outline

- Part I. Data-Driven Problem Solving
- Part II. Past and Current Projects
  - Scholarly Research Output in Zambia
  - Predicting Learning Outcome at UNZA
  - Medical Imaging Workflows in Zambia
  - Automatic Weather Prediction in Zambia
- Part III. Potential Problems
Outline

● Part I. Data-Driven Problem Solving

● Part II. Past and Current Projects
  ○ **Scholarly Research Output in Zambia**
  ○ Predicting Learning Outcome at UNZA
  ○ Medical Imaging Workflows in Zambia
  ○ Automatic Weather Prediction in Zambia

● Part III. Potential Problems
Project #1: Online Visibility of Research in Zambia—Problem (1/4)

https://worldmapper.org
Project #1: Online Visibility of Research in Zambia—Problem (1/4)

https://worldmapper.org
Project #1: Online Visibility of Research in Zambia—Problem (2/4)
Project #1: Online Visibility of Research in Zambia—Problem (2/4)
### Project #1: Online Visibility of Research in Zambia—Problem (3/4)

<table>
<thead>
<tr>
<th>ranking</th>
<th>World Rank</th>
<th>University</th>
<th>Det.</th>
<th>Presence Rank*</th>
<th>Impact Rank*</th>
<th>Openness Rank*</th>
<th>Excellence Rank*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2229</td>
<td>University of Zambia</td>
<td></td>
<td>2948</td>
<td>5899</td>
<td>1855</td>
<td>1876</td>
</tr>
<tr>
<td>2</td>
<td>5073</td>
<td>Copperbelt University</td>
<td></td>
<td>4222</td>
<td>13286</td>
<td>5276</td>
<td>4215</td>
</tr>
<tr>
<td>3</td>
<td>11148</td>
<td>Mulungushi University</td>
<td></td>
<td>4500</td>
<td>16531</td>
<td>6103</td>
<td>5624</td>
</tr>
<tr>
<td>4</td>
<td>14524</td>
<td>Zambian Open University</td>
<td></td>
<td>20785</td>
<td>15925</td>
<td>7356</td>
<td>5624</td>
</tr>
<tr>
<td>5</td>
<td>15556</td>
<td>University of Lusaka</td>
<td></td>
<td>6610</td>
<td>15468</td>
<td>7356</td>
<td>6084</td>
</tr>
<tr>
<td>6</td>
<td>15649</td>
<td>Cavendish University Zambia</td>
<td></td>
<td>8759</td>
<td>15384</td>
<td>7356</td>
<td>6084</td>
</tr>
<tr>
<td>7</td>
<td>17496</td>
<td>Kwame Nkrumah University</td>
<td></td>
<td>9810</td>
<td>17568</td>
<td>7356</td>
<td>6084</td>
</tr>
<tr>
<td>8</td>
<td>17700</td>
<td>Northrise University</td>
<td></td>
<td>12878</td>
<td>17554</td>
<td>7356</td>
<td>6084</td>
</tr>
<tr>
<td>9</td>
<td>18706</td>
<td>ZCAS University</td>
<td></td>
<td>14080</td>
<td>18587</td>
<td>7356</td>
<td>6084</td>
</tr>
<tr>
<td>10</td>
<td>19855</td>
<td>African Christian University</td>
<td></td>
<td>24372</td>
<td>18943</td>
<td>7356</td>
<td>6084</td>
</tr>
<tr>
<td>11</td>
<td>20764</td>
<td>Rusangu University (Zambia Adventist University)</td>
<td></td>
<td>23350</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

http://www.webometrics.info
Project #1: Online Visibility of Research in Zambia—Problem (3/4)

<table>
<thead>
<tr>
<th>ranking</th>
<th>World Rank</th>
<th>University</th>
<th>Det. Rank*</th>
<th>Presence Rank*</th>
<th>Impact Rank*</th>
<th>Openness Rank*</th>
<th>Excellence Rank*</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2229</td>
<td>University of Zambia</td>
<td>2948</td>
<td>5899</td>
<td>1855</td>
<td>1876</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5073</td>
<td>Copperbelt University</td>
<td>4222</td>
<td>13286</td>
<td>5276</td>
<td>4215</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11148</td>
<td>Mulungushi University</td>
<td>4500</td>
<td>16531</td>
<td>6103</td>
<td>5624</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>14524</td>
<td>Zambian Open University</td>
<td>20785</td>
<td>15925</td>
<td>7356</td>
<td>5624</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>15556</td>
<td>University of Lusaka</td>
<td>6610</td>
<td>15468</td>
<td>7356</td>
<td>6084</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>15649</td>
<td>Cavendish University Zambia</td>
<td>8759</td>
<td>15384</td>
<td>7356</td>
<td>6084</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>17496</td>
<td>Kwame Nkrumah University</td>
<td>9810</td>
<td>17568</td>
<td>7356</td>
<td>6084</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>17700</td>
<td>Northrise University</td>
<td>12878</td>
<td>17554</td>
<td>7356</td>
<td>6084</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>18706</td>
<td>ZCAS University</td>
<td>14080</td>
<td>18587</td>
<td>7356</td>
<td>6084</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>19855</td>
<td>African Christian University</td>
<td>24372</td>
<td>18943</td>
<td>7356</td>
<td>6084</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>20764</td>
<td>Rusangu University (Zambia Adventist University)</td>
<td>23350</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

http://www.webometrics.info
Scholarly research and publication form an integral part of the core functions of Higher Education Institutions (HEIs). It is generally standard practice for HEIs to deposit scholarly output into publicly accessible Institutional Repositories (IRs). While Zambia has seen a rise in the number of HEIs, with a total of six Public HEIs and 60 Private HEIs, there is little online visibility of scholarly output generated by these HEIs. A bibliometric analysis, focused on electronic theses and dissertations (ETDs), was conducted by harvesting scholarly publications from HEIs IRs, in order to demonstrate the low online visibility of scholarly research output in Zambia. We also outline technological initiatives, by using case examples from The University of Zambia, that can be employed to potentially increase the online visibility of HEIs scholarly output. Specifically, we illustrate...
Project #1: Online Visibility of Research in Zambia—Problem (4/4)

Research Visibility in the Global South: Towards Increased Online Visibility of Scholarly Research Output in Zambia

and publication form an integral part of the core functions of Higher Education Institutions (HEIs). It is generally standard practice for HEIs to deposit their scholarly publications to publicly accessible Institutional Repositories (IRs). While Zambia has a modest number of HEIs, with a total of six Public HEIs and 60 Private HEIs, the visibility of scholarly output generated by these HEIs is low. A bibliometric analysis of electronic theses and dissertations (ETDs), was conducted by Phiri et al. (2018) on Google Scholarly publications from HEIs IRs, in order to demonstrate the low online visibility of scholarly output in Zambia. We also outline technological initiatives, such as The University of Zambia’s Library and the UZIR, that can be employed to improve the online visibility of HEIs scholarly output. Specifically, we illustrate that...

Phiri, L. (2018)
“Towards Increased Online Visibility of Scholarly Research Output in Zambia”.
URL: http://lis.unza.zm/archive/handle/123456789/227
Project #1: Online Visibility of Research in Zambia—Problem (4/4)

Research Visibility in the Global South: Towards Increased Online Visibility of Scholarly Research Output

Phiri, L. (2018)
“Towards Increased Online Visibility of Scholarly Research Output in Zambia”. URL: http://lis.unza.zm/archive/handle/123456789/227
Project #1: Online Visibility of Research in Zambia—Multipronged Approach

Open Access Electronic Publishing for Increased Online Visibility: Tooling Challenges and Potential Solutions

Lighton Phiri <lighton.phiri@unza.zm>
Department of Library and Information Science
University of Zambia

ZCAS Colloquium on “Sharing Knowledge and Best Practices”
Project #1: Online Visibility of Research in Zambia—Multipronged Approach

Towards Open Access Institutional Repositories in Zambian HEIs

Abel M’kulama <abel.mkulama@unza.zm>
Lighton Phiri <lighton.phiri@unza.zm>

Department of Library and Information Science
University of Zambia

ZCAS Colloquium on “Sharing Knowledge and Best Practices”
Project #1: Online Visibility of Research in Zambia—Multipronged Approach

Automatic classification of digital objects for improved metadata quality of electronic theses and dissertations in institutional repositories

Lighton Phiri

https://doi.org/10.1504/IJMSO.2020.112804

Published online 26 January 2021
Automatic classification of digital objects for improved metadata quality of electronic theses and dissertations in institutional repositories

Lighton Phiri

https://doi.org/10.1504/IJMSO.2020.112804

Published online 26 January 2021
Project #1: Online Visibility of Research in Zambia—ETDs Automatic Classification (1/7)

- Implementation of classification models to automatically classify IR digital objects using the minimum possible input from graduate students: “The ETD Manuscript”
  - The ETD manuscript bitstream is considered the “single source of truth”
  - Metadata prepared by staff that work with IR potentially have inconsistencies

Phiri, L. (2021)
“Automatic Classification of Digital Objects for Improved Metadata Quality of ETDs”
URL: https://doi.org/10.1504/IJMSO.2020.112804
● Text features extracted from a set of core bitstream portions—ETD Title, ETD Abstract, ETD Title Page and ETD pages—to classify ETD manuscripts
Project #1: Online Visibility of Research in Zambia—ETDs Automatic Classification (3/7)

- Textual content mined from PDF manuscripts
  - Cover/title pages
  - Preliminary pages
- Textual content mined from metadata for training
- PDF document metadata
- Curated datasets from external repositories
Project #1: Online Visibility of Research in Zambia—ETDs Automatic Classification (3/7)

- Textual content mined from PDF manuscripts
  - Cover/title pages
  - Preliminary pages
- Textual content mined from metadata for training
- PDF document metadata
- Curated datasets from external repositories
Project #1: Online Visibility of Research in Zambia—ETDs Automatic Classification (3/7)

- Textual content mined from PDF manuscripts
  - Cover/title pages
  - Preliminary pages
- Textual content mined from metadata for training
- PDF document metadata
- Curated datasets from external repositories
Project #1: Online Visibility of Research in Zambia—ETDs Automatic Classification (3/7)

- Textual content mined from PDF manuscripts
  - Cover/title pages
  - Preliminary pages
- Textual content mined from metadata for training
- PDF document metadata
- Curated datasets from external repositories
Project #1: Online Visibility of Research in Zambia—ETDs Automatic Classification (4/7)

- OAI-PMH used to harvest all ETD descriptive metadata elements
- OAI-ORE used to harvest all ETD PDF documents

Metadata

```xml
<atom:source>
  <atom:generator>The University of Zambia Institutional Repository</atom:generator>
  <atom:source>
  <atom:title>Knowledge, attitude and practice of cholera outbreaks in nchelenge district.</atom:title>
  <atom:author>
    <atom:name>Kabita, Kabita.</atom:name>
  </atom:author>
  <atom:category label="DSpace Item" term="DSpaceItem" scheme="http://www.dspace.org/objectModel"/>
  <ore:atom:triples>
    <dc:date>2018</dc:date>
    <dc:type>Article</dc:type>
    <dc:identifier>ISSN: 3471-7102</dc:identifier>
  </atom:link>
</atom:source>
```

May 25, 2021
Project #1: Online Visibility of Research in Zambia—ETDs Automatic Classification (5/7)

- ETD Type—98.1%
- ETD Collection—81.1%
- ETD Subjects—81.7%
- The models would still need to be incorporated into an application that requires “some” human intervention
A Flask API that processes POST requests to classify ETDs

https://github.com/lightonphiri/etd_autoclassifier
Project #1: Online Visibility of Research in Zambia—ETDs Automatic Classification (7/7)

```json
{
    "collectionPrediction": {
        "collectionCode": 0,
        "collectionName": "Education"
    }
}
```

### POST

```
https://datalab-apis.herokuapp.com/api/collection
```

<table>
<thead>
<tr>
<th>Params</th>
<th>Authorization</th>
<th>Headers (9)</th>
<th>Body</th>
<th>Pre-request Script</th>
<th>Tests</th>
<th>Settings</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>form-data</td>
<td>x-www-form-urlencoded</td>
<td>raw</td>
<td>binary</td>
<td>GraphQL BETA</td>
<td>Text</td>
</tr>
</tbody>
</table>

1. "title": "LECTURERS’ PREPAREDNESS TO TRAIN TEACHERS OF LITERACY AND LANGUAGE EDUCATION IN COLLEGES OF EDUCATION IN ZAMBIA", "abstract": "The issue of literacy in Zambia has
pupils. Literacy should be the main focus if any improvement in the education system is to be done. The focus of literacy education is to improve the reading and writing
study was to establish whether literacy and language lecturers were adequately prepared to train teachers in literacy and language education in colleges of education in a
lecturers to prepare teacher trainees in literacy and language education, establish whether the literacy and language teacher education programme adequately responds to a
language lecturers face in preparing trainee teachers in literacy and language education. The study employed a mixed method descriptive design which involved both qualitative
with 49 respondents, that is, 45 college lecturers and four college administrators while simple random sampling was used to select colleges of education where data was col-
that while all literacy and language lecturers were professionally trained either as primary school teachers or secondary school teachers, they were not fully prepared to
interpret the literacy and language teacher education programme. The literacy and language teacher education curriculum not being totally in line with the school curriculum
update literacy and language lecturers on the latest developments of the school curriculum and shortage of lecturers was among challenges. The study recommended that let-
to prepare them to train teachers, the literacy and language curriculum should always be updated in line with the school curriculum in order for colleges of education to
lectures in the formulation and /or revision of literacy policies.";

https://datalab-apis.herokuapp.com/api/collection
# Project #1: Online Visibility of Research in Zambia—Current Work (1/3)

<table>
<thead>
<tr>
<th>Classifier Chains</th>
<th>Title</th>
<th>Abstract</th>
<th>Title + Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>F1 Score</td>
<td>Hamming Loss</td>
<td>Jaccard Score</td>
</tr>
<tr>
<td>MultinomialNB</td>
<td>TF</td>
<td>0.060</td>
<td>0.055</td>
</tr>
<tr>
<td></td>
<td>TF-IDF</td>
<td>0.090</td>
<td>0.027</td>
</tr>
<tr>
<td>RandomForest</td>
<td>TF</td>
<td>0.287</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>TF-IDF</td>
<td>0.289</td>
<td>0.009</td>
</tr>
<tr>
<td>SGDClassifier</td>
<td>TF</td>
<td>0.312</td>
<td>0.006</td>
</tr>
<tr>
<td></td>
<td>TF-IDF</td>
<td>0.310</td>
<td>0.006</td>
</tr>
</tbody>
</table>

## One-Versus-Rest

<table>
<thead>
<tr>
<th>Title</th>
<th>Abstract</th>
<th>Title + Abstract</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 Score</td>
<td>Hamming Loss</td>
<td>Jaccard Score</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*M’sendo R. (2019—Present)*  
*MSc Computer Science, University of Zambia*  
“Multi-Faceted Automatic Classification of Institutional Repository Objects”
3. Metadata Elements

The following is a description of the common Dublin Core metadata elements (and a new element specifically for theses). Guidelines are given as to which information related to an ETD belongs in each element.

If a more general element is described as mandatory, it should be specified fully, even if qualified elements are also specified. If a more general element is repeatable, it should be assumed that sub-elements are also repeatable. If an element contains free text, it must be repeatable to allow for ETDs that provide metadata in more than one language.

### 3.1 dc.title

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>dc.title</td>
<td>A name given to the resource. In the case of theses and dissertations, this is the title of the work as it appears on the title page or equivalent.</td>
<td>Mandatory, Repeatable</td>
</tr>
<tr>
<td>dc.title.alternative</td>
<td>alternative title of the thesis or dissertation</td>
<td>Optional, Repeatable</td>
</tr>
</tbody>
</table>

### 3.2 dc.creator

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
<th>Notes</th>
</tr>
</thead>
</table>

Chisale A. (2021—Present)
MLIS, University of Zambia
“Automatic Generation of Electronic Theses and Dissertations Metadata”
Project #1: Online Visibility of Research in Zambia—Current Work (3/3)

Recent Submissions

1. Theory and policy in export credit insurance and finance: an analytical and empirical study
   
   Tue, 11 Dec 2018 18:03:43 UTC

2. Land tenure and agricultural development in Zambia
   
   Tue, 11 Dec 2018 18:03:43 UTC

3. State, trade unions and labour policies: theoretical issues and evidence from Zambia
   
   Tue, 11 Dec 2018 18:03:43 UTC

4. Export promotion possibilities in the industrial sector of Zambia: a neo-classical approach
   
   Tue, 11 Dec 2018 18:03:43 UTC

5. The 1980 decentralisation and district level planning in Zambia
   
   Tue, 11 Dec 2018 18:03:43 UTC
Outline

● Part I. Data-Driven Problem Solving

● Part II. Past and Current Projects
  ○ Scholarly Research Output in Zambia
  ○ Predicting Learning Outcome at UNZA
  ○ Medical Imaging Workflows in Zambia
  ○ Automatic Weather Prediction in Zambia

● Part III. Potential Problems
Project #2: Predicting Student Learning Outcomes—Problem (1/2)

- ICT 1110 performance is an issue. The poor performance transcends all assessments: quizzes, tests and practical programming questions.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td>1</td>
<td>3</td>
<td>20</td>
<td>3.5</td>
<td>5.5</td>
<td>10</td>
<td>2</td>
<td>24</td>
<td>3.85</td>
<td>6</td>
<td>4</td>
<td>22</td>
<td>0.5</td>
<td>5</td>
<td>7</td>
<td>31.5</td>
<td>0</td>
<td>7</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>10</td>
<td>14.6</td>
<td>11.9</td>
<td>26.5</td>
<td>53</td>
<td>11</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angola</td>
<td>5</td>
<td>3</td>
<td>22.5</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>6</td>
<td>17</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>17</td>
<td>1</td>
<td>7</td>
<td>9</td>
<td>20.5</td>
<td>0</td>
<td>2.5</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>14.6</td>
<td>11.6</td>
<td>13.3</td>
<td>24.9</td>
<td>49.8</td>
<td>18</td>
<td>40.2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benin</td>
<td>4</td>
<td>2.5</td>
<td>20.5</td>
<td>4.5</td>
<td>5</td>
<td>9</td>
<td>1</td>
<td>12.5</td>
<td>5</td>
<td>5.5</td>
<td>0</td>
<td>3</td>
<td>22</td>
<td>10</td>
<td>0.5</td>
<td>5</td>
<td>9</td>
<td>22.5</td>
<td>34</td>
<td>1</td>
<td>8</td>
<td>10</td>
<td>8</td>
<td>9</td>
<td>14.9</td>
<td>10.3</td>
<td>22.5</td>
<td>45</td>
<td>24</td>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Botswana</td>
<td>3</td>
<td>2</td>
<td>16</td>
<td>5</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>21</td>
<td>5</td>
<td>9</td>
<td>0</td>
<td>1</td>
<td>14</td>
<td>10</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>26</td>
<td>26</td>
<td>1</td>
<td>0</td>
<td>10</td>
<td>8</td>
<td>9</td>
<td>13.2</td>
<td>8</td>
<td>21.2</td>
<td>42.4</td>
<td>37</td>
<td>47.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>6.5</td>
<td>4</td>
<td>17</td>
<td>5.5</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>20.5</td>
<td>7</td>
<td>7</td>
<td>1.5</td>
<td>3</td>
<td>19</td>
<td>7</td>
<td>0.5</td>
<td>4</td>
<td>6</td>
<td>17.5</td>
<td>27</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>8</td>
<td>9</td>
<td>12.6</td>
<td>8.3</td>
<td>20.9</td>
<td>41.8</td>
<td>40</td>
<td>48.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burundi</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>2.5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>17</td>
<td>4</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>15</td>
<td>10</td>
<td>1</td>
<td>6</td>
<td>6</td>
<td>21</td>
<td>24</td>
<td>8</td>
<td>9</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>11.6</td>
<td>10</td>
<td>21.6</td>
<td>43.2</td>
<td>35</td>
<td>46.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Project #2: Predicting Student Learning Outcomes—Problem (1/2)

- ICT 1110 performance is an issue. The poor performance transcends all assessments: quizzes, tests, and practical programming questions.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Algeria</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Angola</td>
<td>5</td>
<td>3</td>
<td>22.5</td>
<td>5</td>
<td>9</td>
<td>6</td>
<td>17</td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>5</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Benin</td>
<td>4</td>
<td>2.5</td>
<td>20.5</td>
<td>4.5</td>
<td>9</td>
<td>1</td>
<td>12.5</td>
<td>5</td>
<td>5.5</td>
<td>0</td>
<td>3</td>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Botswana</td>
<td>3</td>
<td>2</td>
<td>16</td>
<td>5</td>
<td>4</td>
<td>8</td>
<td>1</td>
<td>21</td>
<td>5</td>
<td>9</td>
<td>0</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burkina Faso</td>
<td>6.5</td>
<td>4</td>
<td>17</td>
<td>5.5</td>
<td>6</td>
<td>4</td>
<td>20.5</td>
<td>7</td>
<td>7</td>
<td>1.5</td>
<td>3</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Burundi</td>
<td>0</td>
<td>1</td>
<td>10</td>
<td>2.5</td>
<td>4</td>
<td>5</td>
<td>4</td>
<td>17</td>
<td>4</td>
<td>8</td>
<td>3</td>
<td>2</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Project #2: Predicting Student Learning Outcomes—Problem (2/2)

- Potential solution: implement a prediction model aimed at identifying at-risk students.
  - Initiate interventions on at-risk students.
Project #2: Predicting Student Learning Outcomes—Data Sources (1/5)

- Demographics information
- LMS interaction logs
- Course workload
- Subject responses
Project #2: Predicting Student Learning Outcomes—Data Sources (2/5)

- Assessment results broken down by question
  - Concepts associated with question
  - Topics associated with question
Project #2: Predicting Student Learning Outcomes—Data Sources (3/5)

- Assessment results broken down by question
  - Concepts associated with question
  - Topics associated with question
Project #2: Predicting Student Learning Outcomes—Data Sources (4/5)

- **LMS interaction logs**
  - How often do students access Moodle (login attempts)
  - Which Moodle features are being accessed (GradeBook, Messaging)
  - Time spent on Moodle

---

May 25, 2021
Project #2: Predicting Student Learning Outcomes—Data Sources (5/5)

- ICT 1110 information survey to capture information not available in SIS
  - Experience with computers
  - Motivation for taking the course
  - Specific location where student lives (although this can be inferred from next of kin address perhaps?)
# Project #2: Predicting Student Learning Outcomes—Current Work


### Performance Predictor: Machine Learning Tool for Student Performance Outcomes

<table>
<thead>
<tr>
<th>Student ID</th>
<th>Algeria</th>
<th>Angola</th>
<th>Benin</th>
<th>Botswana</th>
<th>Burkina Faso</th>
<th>Burundi</th>
<th>Cameroon</th>
<th>Cape Verde</th>
<th>Central African</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quiz #2</td>
<td>1</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>6.5</td>
<td>0</td>
<td>3.3</td>
<td>4.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Quiz #3</td>
<td>3</td>
<td>22.5</td>
<td>9.1</td>
<td>2.1</td>
<td>14.0</td>
<td>1.0</td>
<td>6.9</td>
<td>8.4</td>
<td>0.0</td>
</tr>
<tr>
<td>Class Test #1</td>
<td>20.3</td>
<td>3.5</td>
<td>5.5</td>
<td>5.5</td>
<td>17.5</td>
<td>7.7</td>
<td>4.5</td>
<td>19.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Class Test #2</td>
<td>5.5</td>
<td>10.0</td>
<td>12.5</td>
<td>6.5</td>
<td>9.5</td>
<td>3.0</td>
<td>15.0</td>
<td>4.5</td>
<td>0.0</td>
</tr>
<tr>
<td>Class Test #3</td>
<td>8.5</td>
<td>6.0</td>
<td>3.5</td>
<td>5</td>
<td>7.5</td>
<td>3.0</td>
<td>15.0</td>
<td>2</td>
<td>0.0</td>
</tr>
<tr>
<td>Class Test #4</td>
<td>22.5</td>
<td>10.0</td>
<td>10.0</td>
<td>10.0</td>
<td>17.5</td>
<td>8.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Class Test #5</td>
<td>10.0</td>
<td>5.5</td>
<td>5.0</td>
<td>5</td>
<td>10.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Class Test #6</td>
<td>5.0</td>
<td>22.5</td>
<td>5.5</td>
<td>4</td>
<td>5.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Class Test #7</td>
<td>9.0</td>
<td>5.0</td>
<td>6</td>
<td>9.0</td>
<td>3.0</td>
<td>2.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Class Test #8</td>
<td>10.0</td>
<td>5</td>
<td>5.0</td>
<td>5</td>
<td>10.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Class Test #9</td>
<td>5.0</td>
<td>22.5</td>
<td>5.5</td>
<td>4</td>
<td>5.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Class Test #10</td>
<td>10.0</td>
<td>5.0</td>
<td>6</td>
<td>9.0</td>
<td>3.0</td>
<td>2.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Class Test #11</td>
<td>5.0</td>
<td>22.5</td>
<td>5.5</td>
<td>4</td>
<td>5.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Class Test #12</td>
<td>10.0</td>
<td>5</td>
<td>5.0</td>
<td>5</td>
<td>10.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Quiz #11</td>
<td>10.0</td>
<td>5.0</td>
<td>6</td>
<td>9.0</td>
<td>3.0</td>
<td>2.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Quiz #12</td>
<td>5.0</td>
<td>22.5</td>
<td>5.5</td>
<td>4</td>
<td>5.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Quiz #13</td>
<td>10.0</td>
<td>5</td>
<td>5.0</td>
<td>5</td>
<td>10.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Quiz #14</td>
<td>5.0</td>
<td>22.5</td>
<td>5.5</td>
<td>4</td>
<td>5.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Quiz #15</td>
<td>10.0</td>
<td>5</td>
<td>5.0</td>
<td>5</td>
<td>10.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Quiz #16</td>
<td>5.0</td>
<td>22.5</td>
<td>5.5</td>
<td>4</td>
<td>5.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Quiz #17</td>
<td>10.0</td>
<td>5</td>
<td>5.0</td>
<td>5</td>
<td>10.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Quiz #18</td>
<td>5.0</td>
<td>22.5</td>
<td>5.5</td>
<td>4</td>
<td>5.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Quiz #19</td>
<td>10.0</td>
<td>5</td>
<td>5.0</td>
<td>5</td>
<td>10.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Quiz #20</td>
<td>5.0</td>
<td>22.5</td>
<td>5.5</td>
<td>4</td>
<td>5.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Make-up Test</td>
<td>10.0</td>
<td>5</td>
<td>5.0</td>
<td>5</td>
<td>10.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>Test Weight</td>
<td>10.0</td>
<td>5</td>
<td>5.0</td>
<td>5</td>
<td>10.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>CA [N/50]</td>
<td>10.0</td>
<td>5</td>
<td>5.0</td>
<td>5</td>
<td>10.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>CA [% Score]</td>
<td>10.0</td>
<td>5</td>
<td>5.0</td>
<td>5</td>
<td>10.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
<tr>
<td>CA [RANK in Class]</td>
<td>10.0</td>
<td>5</td>
<td>5.0</td>
<td>5</td>
<td>10.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

### Exam % to Pass

- Algeria: 25%
- Angola: 20%
- Benin: 25%
- Botswana: 25%
- Burkina Faso: 25%
- Burundi: 25%
- Cameroon: 25%
- Cape Verde: 25%
- Central African: 25%
Outline

● Part I. Data-Driven Problem Solving
● Part II. Past and Current Projects
   ○ Scholarly Research Output in Zambia
   ○ Predicting Learning Outcome at UNZA
   ○ Medical Imaging Workflows in Zambia
   ○ Automatic Weather Prediction in Zambia
● Part III. Potential Problems
Project #3: Medical Imaging Workflows in Zambia—Problem

Need for Image Reporting by Radiographers in Zambia

Osward Bwanga
Midland Regional Hospital at Tullamore

Joseph Mulenga
Warwick Hospital Warwick, UK

Ernest Chanda
Cancer Diseases Hospital, Zambia

Joseph Mulenga
Warwick Hospital

Keywords: Reporting radiographer, radiologist, role extension, plain film radiography

https://mjz.co.zm/index.php/mjz/article/view/560
Project #3: Medical Imaging Workflows in Zambia—Current Work (1/2)
Project #3: Medical Imaging Workflows in Zambia—Current Work (1/2)
Project #3: Medical Imaging Workflows in Zambia—Current Work (2/2)
Project #3: Medical Imaging Workflows in Zambia—Current Work (2/2)
Project #3: Medical Imaging Workflows in Zambia—Current Work (2/2)
Outline

• Part I. Data-Driven Problem Solving
• Part II. Past and Current Projects
  ○ Scholarly Research Output in Zambia
  ○ Predicting Learning Outcome at UNZA
  ○ Medical Imaging Workflows in Zambia
  ○ Automatic Weather Prediction in Zambia
• Part III. Potential Problems
Project #3: Automatic Forecasting of Seasonal Rainfall—Current Work

Figure 1: Rainfall Forecast for October, November & December 2020 (OND).

Figure 2: Rainfall Forecast for November, December 2020 & January 2021 (NDJ).
Outline

- Part I. Data-Driven Problem Solving
- Part II. Past and Current Projects
- Part III. Potential Problems
  - Exemplar Projects in Zambia
  - Potential Locally Relevant Problems
Outline

- Part I. Data-Driven Problem Solving
- Part II. Past and Current Projects
- Part III. Potential Problems
  - Exemplar Projects in Zambia
  - Potential Locally Relevant Problems
Agriculture: Automatic identification and Early Warning of Fall Armyworms

Developing an automated fall armyworm identification, early warning and monitoring system using a convolution neural network

Since its reported presence in 2016, the Fall Armyworm (FAW) has caused major damage to a number of plant species including maize which is a staple food for most African countries. Their presence in Africa poses a challenge to the food security in many countries contributing to the already existing food problem that the continent has been facing. The outbreak of FAW if left uncontrolled is likely to have a devastating impact on the food supply. Zambia has not been spared by this scourge and currently both commercial and subsistence farmers in Zambia use a manual process to monitor crops for the presence of the fall army worm. Manual monitoring and observation is time consuming, labour intensive and also results in the delay of applying appropriate pest control measures by farmers. There is need to improve on the methods used in the monitoring of the pest if there is to be a proactive and reactive response by stakeholders to FAW. This study was therefore aimed at improving on the FAW pest monitoring and early warning based on ...
Telecommunications: Automatic Customer Segmentation

Product recommender system for telecommunication industries: a case of Zambia telecommunications companies

Recommender systems have become increasingly popular in recent years, and are utilized in a variety of areas including movies, music, news, products, research articles, search queries, social tags, and products in general they are designed to automatically generate personalized suggestions of products/services to customers. With the competitiveness that is growing in the telecommunication industry, telecommunication operators seek ways to attract and keep the subscribers on their network. It’s notable that telecommunication operators lack the ability to manage their customer retention rate because they do not have a personalized way of recommending products and services to their subscribers, as a result subscribers tend to migrate to new providers. This trend of subscribers migrating to new providers proves to be a severe problem for Telecommunication providers as they experience subscriber base and revenue shrinkage. This dissertation describes a Recommender System for Telecommunication companies using call detail reports (CDR's), machine learning algorithms and big data concepts. Experimental results are presented.
A Data Mining Model for Predicting and Forecasting Fraud in Banks

Knox Kamusweke
Dept. of Computer Science
University of Zambia
P.O. Box 32379, Lusaka, Zambia
Email: knox.kamusweke@cs.unza.zm

Mayumbo Nyirenda
Dept. of Computer Science
University of Zambia
P.O. Box 32379, Lusaka, Zambia
Email: mayumbo.nyirenda@cs.unza.zm

Monde Kabemba
Dept. of Computer Science
University of Zambia
P.O. Box 32379, Lusaka, Zambia
Email: monde.kalumbilo@cs.unza.zm

Abstract — Banks generate and handle millions of transactions across their platforms. These transactions contain significant patterns and trends which are hidden but needed for knowledge discovery and actionable insight. Uncovering these patterns and trends has always been a challenge for most financial institutions due to the large volumes of transactions and the ever changing patterns. This has made the patterns and trends more sophisticated to discover and has introduced a need for more robust automated systems that can recognize some temporal characteristics and patterns that have to be identified.

The Reserve Bank of India – RBI maintains data on frauds on the basis of area of operation under which the frauds have been perpetrated [4]. According to such data pertaining to top ten categories under which frauds have been...
Outline

- Part I. Data-Driven Problem Solving
- Part II. Past and Current Projects
- Part III. Potential Problems
  - Exemplar Projects in Zambia
  - Potential Locally Relevant Problems
Potential Locally Relevant Problems in Zambia (1/6)

- Impact-driven research/studies
  - Education
  - Health
  - So-called ICT for development perhaps?
Potential Locally Relevant Problems in Zambia (2/6)

● Impact-driven research/studies
  ○ Education
  ○ Health
  ○ So-called ICT for development perhaps?
Potential Locally Relevant Problems in Zambia (3/6)

- Impact-driven research/studies
  - Education
  - Health
  - So-called ICT for development perhaps?
Potential Locally Relevant Problems in Zambia (3/6)

- Impact-driven research/studies
  - Education
  - Health
  - So-called ICT for development perhaps
Potential Locally Relevant Problems in Zambia (4/6)

● Impact-driven research/studies
  ○ Education
  ○ Health
  ○ So-called ICT for development perhaps?
Potential Locally Relevant Problems in Zambia (4/6)

- Impact-driven research/studies
  - Education
  - Health
  - So-called ICT for development perhaps?
Potential Locally Relevant Problems in Zambia (4/6)

- Impact-driven research/studies
  - Education
  - Health
  - So-called ICT for development perhaps?
Potential Locally Relevant Problems in Zambia (4/6)

- Impact-driven research/studies
  - Education
  - Health
  - So-called ICT for development perhaps?
Potential Locally Relevant Problems in Zambia (5/6)

- Impact-driven research/studies
  - Education
  - Health
  - So-called ICT for development perhaps?
Potential Locally Relevant Problems in Zambia (5/6)

- Impact-driven research/studies
  - Education
  - Health
  - So-called ICT for development perhaps?
Potential Locally Relevant Problems in Zambia (6/6)

- Education
- Health
- So-called ICT for development perhaps?
Potential Locally Relevant Problems in Zambia (6/6)

- Education
- Health
- So-called ICT for development perhaps?
Q & A Session

● Comments, concerns and complaints?
Bibliography


lighton.phiri@unza.zm
http://datalab.unza.zm
http://lis.unza.zm/~lightonphiri