Plan Overview

A Data Management Plan created using DMPTool

DMP ID: https://doi.org/10.48321/D1CC5C9841

Title: The use of data science to improve mathematic performance in rural schools in the North west province

Creator: Motshidisi Khutlapye - **ORCID:** 0009-0005-3386-3384

Affiliation: Iie Varsity College

Contributor: Dr Samwel Mwapwele

Funder: The Independent Institute of Education (iie.ac.za)

Template: Digital Curation Centre

Project abstract:

In the North West Province of South Africa, rural schools face numerous challenges in providing quality mathematics education, leading to persistently low levels of mathematical performance among students. This study explores the potential of data science methodologies to address these challenges and improve mathematical performance in rural schools.

The research adopts quantitative analysis of student performance data with qualitative insights from educators and stakeholders. Quantitative data will be collected from rural schools in the North West Province, encompassing student mathematics scores, attendance records, demographic information, and other relevant variables. Data science techniques, including predictive modeling, machine learning algorithms, and statistical analysis, will be employed to identify patterns, trends, and predictive factors associated with mathematical performance.

Start date: 05-15-2024

End date: 11-30-2024

Last modified: 05-15-2024

Copyright information:

The use of data science to improve mathematic performance in rural schools in the North west province

Data Collection

What data will you collect or create?

Secondary Data and Quantitave type of data.

Collecting numerical data such as student mathematics scores, attendance records, and demographic information.
Applying statistical methods such as regression analysis, correlation analysis, and predictive modeling to identify patterns, trends, and predictive factors associated with mathematical performance.
Quantifying the relationships between variables to understand the impact of factors such as student demographics, teacher qualifications, and classroom resources on mathematical achievement.
Using numerical data and statistical analysis to forecast future performance trends and evaluate the effectiveness of potential interventions.

How will the data be collected or created?

The data will be collected from the Department of Education through the request process public request

Documentation and Metadata

What documentation and metadata will accompany the data?

- Data Dictionary: A comprehensive document that provides descriptions of each variable in the dataset, including its name, type, definition, and possible values. This helps users understand the meaning and interpretation of the data variables.
- Data Collection Protocol: A detailed description of the methods and procedures used to collect the data, including information on sampling techniques, data sources, data collection instruments, and any quality control measures implemented during data collection.
- Codebook or Data Code: A document that contains the code used for data cleaning, processing, and analysis. This includes any scripts, algorithms, or programming code used to transform raw data into usable formats and perform statistical analyses.
- Metadata: Information about the dataset itself, including its title, description, creator, date of creation, version number, and any relevant keywords or tags. Metadata helps users locate, understand, and evaluate the dataset.
- Data Usage Policy: Guidelines and restrictions on the permissible uses of the data, including any privacy or ethical considerations, data sharing agreements, copyright or licensing terms, and data access restrictions.
- Data Cleaning and Preprocessing Logs: Documentation of the steps taken to clean, preprocess, and prepare the data for analysis, including any transformations, imputations, or outlier removal procedures applied to the raw data.
- Ethical Considerations: Documentation of any ethical considerations or approvals obtained for data collection and analysis, including Institutional Review Board (IRB) approvals, informed consent procedures, and compliance with relevant privacy regulations.

Ethics and Legal Compliance

How will you manage any ethical issues?

- Ensure Data Accuracy and Integrity: Take steps to ensure the accuracy, reliability, and integrity of the data used in the analysis. Validate data sources, conduct quality checks, and document data cleaning and preprocessing procedures to ensure the validity of the results. Transparently report any limitations or biases in the data that may affect the interpretation of the findings.
- Promote Transparency and Accountability: Be transparent about the methods, assumptions, and limitations of the data science techniques used in the analysis. Clearly document the data analysis process, including the selection of variables, statistical methods, and modeling techniques employed. Make the research findings and methodologies accessible to stakeholders and the broader community to promote accountability and scrutiny.
- Seek Ethical Oversight and Approval: Obtain ethical approval from relevant institutional review boards (IRBs) or ethics committees before initiating any data collection or analysis activities. Follow ethical guidelines and standards for research conduct, such as the principles outlined in the Belmont Report or the Declaration of Helsinki. Consult with experts in data ethics, education research, and privacy law to ensure compliance with ethical norms and regulations.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

- Acknowledge Sources and Attribution: Provide proper attribution and citation for the secondary data sources used in your research, acknowledging the original creators or copyright holders. Follow established citation practices and include clear references to the data sources in your research outputs, such as publications, reports, or presentations.
- Respect Data Use Agreements: Adhere to any data use agreements or terms of use specified by the data provider or repository. Respect any restrictions on data access, redistribution, or commercial use, and ensure that your use of the data is consistent with the agreed-upon terms.

Storage and Backup

How will the data be stored and backed up during the research?

Data will be saved in the IIE One Drive using the IIE Virtual Machine to analyse the data

How will you manage access and security?

Security is managed thoruhg a password restricition which is only known by myself and not shared with any other person

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

- Raw Data: Raw data collected during the research process, including primary data collected from surveys, experiments, observations, or interviews, should be retained as it forms the

foundation of the research findings. Raw data is essential for replication, validation, and further analysis by other researchers.

- Processed Data: Processed or cleaned data that have been transformed or analyzed for specific research purposes should also be retained, as they represent the intermediate stages of data analysis and interpretation. Processed data may include aggregated data, summary statistics, or derived variables.
- Documentation and Metadata: Comprehensive documentation and metadata describing the research data, including data dictionaries, codebooks, methodology descriptions, and variable definitions, should be retained to provide context and facilitate understanding of the data by other researchers.
- Analysis Scripts and Code: Any scripts, algorithms, or code used for data analysis, cleaning, processing, or visualization should be retained to ensure reproducibility and transparency of the research findings. Sharing analysis scripts allows other researchers to replicate the analyses and verify the results independently.

What is the long-term preservation plan for the dataset?

- Data Format and Standards: Standardize the data format and structure of the dataset to ensure compatibility and interoperability with future software applications and data analysis tools. Choose open and non-proprietary data formats whenever possible to mitigate the risk of format obsolescence.
- Metadata and Documentation: Create comprehensive metadata and documentation for the dataset, including descriptions of data variables, methodologies, data collection procedures, and any relevant contextual information. Document metadata according to established standards and best practices to facilitate data discovery, understanding, and reuse.
- Data Versioning and Revision Control: Implement versioning and revision control mechanisms to track changes to the dataset over time and maintain a record of previous versions. Use version control systems or data management platforms to manage revisions, track lineage, and ensure data integrity.
- Data Preservation Policies: Develop data preservation policies and procedures to govern the long-term management and stewardship of the dataset. Define responsibilities, roles, and workflows for data preservation activities, including data migration, format conversion, and metadata maintenance.
- Data Lifecycle Management: Establish data lifecycle management practices to guide the ongoing maintenance, curation, and preservation of the dataset throughout its lifecycle. Monitor data usage, relevance, and obsolescence over time and implement strategies to ensure the continued usability and value of the dataset.

Data Sharing

How will you share the data?

Data Sharing and Dissemination: Facilitate data sharing and dissemination to maximize the impact and utility of the dataset. Publish the dataset in relevant data repositories, archives, or data portals to make it accessible to the broader research community and comply with funder or publisher mandates for data sharing.

Are any restrictions on data sharing required?

Data Access and Permissions: Establish access controls and permissions to regulate access to the dataset and protect sensitive or confidential information. Define access policies, user roles, and authentication mechanisms to manage data access and ensure compliance with ethical and legal requirements.

Responsibilities and Resources

Who will be responsible for data management?

Supervisor

Myself(Researcher)

Ethics Committee

What resources will you require to deliver your plan?

Software and Tools:

Data Management Software: Tools and software applications for data collection, cleaning, analysis, visualization, and documentation.

Statistical Analysis Software: Software packages for statistical analysis and data visualization, such as R, Python

Version Control Systems: Version control software or platforms for managing revisions, tracking changes, and collaborating on research data and code.
