

**MS in Climate Smart Agriculture**  
**Institute of Climate Smart Agriculture**  
**Patuakhali Science and Technology University, Dumki, Patuakhali**

**Course Profile**

**Course Code:** CSA 5115

**Course Title:** Data Management

**Credit Hour:** 2

**Student Level:** Level- 6, Semester – 2

**Rationale:** Students of this course are expected to learn and understand the major concepts of data management, processing, data governance, and basic data mining techniques with emphasis and consideration on climate smart agriculture (CSA)

**Objectives:** At the end of this course, students should be able to:

- Understand the data management challenges and solutions
- Understand principles of data management
- Gain technical expertise on data collection, storage, processing and representation.
- Acquire knowledge on data interpretation and inference
- Apply these knowledge on domain specific data (e.g. Agronomic/Life science/Business data)

<b>Learning Outcomes</b>	<b>Course Content</b>	<b>Teaching-Learning Strategy</b>	<b>Assessment Strategy</b>
<p>Basic concepts of Data &amp; Information</p> <p>Overview of Data Management</p>	<p><b>Introduction:</b> Definition, differences between data and information, types of data (structured, unstructured) and metadata, data formats, types of data sources in agriculture and CSA</p> <p><b>Data Management:</b> Data management services, data stewardship, research lifecycle (, hypothesis and objective setting, discovery &amp; planning, data collection, data organization and processing, data analysis and representation, data publication and sharing, long-term management of data)</p>	<p>Lecture</p> <p>Reading assignment</p> <p>Presentation</p> <p>QA</p>	<p>-Short Answer</p> <p>-Essay</p> <p>-Checklist</p>

<p>Statistical Learning for Data Management</p> <p>(pre-requisite before diving deep into data management details)</p>	<p><b>Descriptive Statistics:</b> Definition, measures of central tendency (mean, median, mode), measures of asymmetry (skewness, kurtosis), measures of variability (variance, standard deviation, coefficient of variation), measures of relationship between variables (covariance and correlation)</p> <p><b>Inferential Statistics:</b> Definition, Common data distribution types (Bernoulli/ Uniform/ Binomial/ Normal/ Poisson Distribution etc.), Central limit theorem, Hypothesis testing, Confidence intervals, Z-scores.</p>	<p>Lecture</p> <p>Reading</p> <p>Discussion</p> <p>Assignment</p> <p>QA</p>	<p>-Short Answer</p> <p>-True-False</p> <p>-Completion</p> <p>-Checklist</p>
<p>Data Visualization Techniques</p> <p>(pre-requisite before diving deep into data management details)</p>	<p><b>Categorical Data Visualization:</b></p> <p>Frequency distribution tables, bar and column charts, Pie charts, Pareto diagrams, cross tables/contingency tables</p> <p><b>Numerical Data Visualization:</b></p> <p>Frequency distribution tables with intervals, Histograms, Box plots, Scatter plots, Line charts</p>	<p>Lecture</p> <p>Presentation</p> <p>Demonstration</p> <p>QA</p>	<p>-Assignment</p> <p>-Practical Exam</p>
<p>Deep Dive in Data Management</p>	<p><b>Discovery and planning:</b> Data management plans (DMPs), Identify area of interest, stakeholders based on research and business potential, Resource management</p> <p><b>Data Collection:</b> Data collection methods, Survey methods, Questionnaire design, Demonstration on popular data</p>	<p>Lecture</p> <p>Reading</p> <p>Discussion</p> <p>Demonstration</p> <p>Assignment</p> <p>QA</p>	<p>-Short Answer</p> <p>-Essay</p> <p>-Presentation</p> <p>-Completion</p> <p>-Practical Exam</p>

	<p>collection toolkits (ODK, Kobo, Google form etc.)</p> <p><b>Data Organization and Processing:</b> Data file formats, Data sorting&amp; grouping, Data cleaning process, Outlier identification, Missing value imputation techniques, Data transformation strategies (Smoothing, Aggregation, Generalization, Normalization, Discretization, Attribute Construction)</p> <p><b>Data Analysis and Representation:</b> Understanding data distribution, Application of statistical learning on acquired data, Application of visualization techniques for proper data visualization and representation.</p> <p><b>Data Publication and Sharing:</b> Data documentation levels, Data citation Data privacy, Data licensing, Data sharing protocols, Data access levels, Embargos, Technological access restrictions, Data use agreements.</p> <p><b>Long-term Management:</b> Data storage and storage facility, Data encryption, Data versioning, Data backup, Cloud storage management, Data modularity.</p>		
<p>More details on Outlier Detection Techniques:</p>	<p><b>Outlier Detection Methods:</b> Supervised, Semi-supervised, Unsupervised methods, Statistical methods, Mahalanobis distance, Box and whisker method etc.</p>	<p>Lecture Reading Discussion Demonstration</p>	<p>-Problem solving -Assignment -Completion</p>

	<p><b>Outlier Replacement Methods:</b> Mean/Median/Mode replacement, Model based replacement (knn, glm, mice etc)</p>	<p>Assignment QA</p>	
<p>Basic Concepts of Data Mining</p>	<p><b>Overview:</b> Data mining definition, Types of data pattern problems (Regression, Classification, Clustering), Parametric, Non-parametric methods.</p> <p><b>Regression:</b> Linear regression, Non-linear regression, Interpolation and extrapolation, Polynomial regression, Bayesian regression</p> <p><b>Classification:</b> Logistic regression, Decision tree, Bayes classifier, Support vector machine</p> <p><b>Clustering:</b> k-means, k-nearest neighbors, Probabilistic hierarchical clustering, DBSCAN</p>	<p>Lecture Reading assignment Discussion QA</p>	<p>-Short Answer -True-False -Problem solving</p>
<p>Data Management in Agricultural and Agronomic domain</p>	<p><b>Overview:</b> Scopes and stakeholder search, Structuring and planning projects, Ways to gather data (direct or passive), Survey design for direct data collection, Finding data source for passive data collection, Procedures of data collection, validation and quality check, application of gathered knowledge of data lifecycles on the acquired data</p>	<p>Lecture Group assignment Demonstration QA</p>	<p>-Completion -Project -Peer rating -Presentation</p>

### Recommended Books and Resources

1. Multi-Domain Master Data Management: Advanced MDM and Data Governance in Practice  
(<https://www.elsevier.com/books/multi-domain-master-data-management/allen/978-0-12-800835-5>)

2. Research Data Management and Sharing  
(<https://www.coursera.org/learn/data-management>)
3. Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance  
(<https://www.elsevier.com/books/data-stewardship/plotkin/978-0-12-822132-7>)
4. Data Mining: Concepts and Techniques  
<http://myweb.sabanciuniv.edu/rdehkharghani/files/2016/02/The-Morgan-Kaufmann-Series-in-Data-Management-Systems-Jiawei-Han-Micheline-Kamber-Jian-Pei-Data-Mining.-Concepts-and-Techniques-3rd-Edition-Morgan-Kaufmann-2011.pdf>
5. Think Stats - Probability and Statistics for Programmers  
(<http://greenteapress.com/thinkstats/thinkstats.pdf>)
6. Statistics and probability  
(<https://www.khanacademy.org/math/statistics-probability>)
7. Pattern Recognition and Machine Learning  
(<https://www.springer.com/gp/book/9780387310732>)
8. Data Visualization with R  
(<https://www.analyticsvidhya.com/blog/2015/07/guide-data-visualization-r/>)
9. Data Visualization with Python  
(<https://www.coursera.org/learn/python-for-data-visualization>)
10. Data Visualization in R  
(<https://www.dataquest.io/course/r-data-viz/>)