EMBA Program
MB-511

Data Science for Managerial Decisions (MB 511)

Instructor
Anant Prakash Awasthi

Data Science for Managerial Decisions (MB 511)

EMBA Program
MB-511

References/Literature

- Jojo Moolayil, “Smarter Decisions The Intersection of IoT and Data Science”, PACKT, 2016.
- Cathy O’Neil and Rachel Schutt , “Doing Data Science”, O'Reilly, 2015.
- David Dietrich, Barry Heller, Beibei Yang, “Data Science and Big data Analytics”, EMC 2013
- Raj, Pethuru, “Handbook of Research on Cloud Infrastructures for Big Data Analytics”, IGI Global
- Management Information System, W.S Jawadekar, Tata Mc Graw Hill Publication.
- Management Information System, David Kroenke, Tata Mc Graw Hill Publication.
- MIS Management Perspective, D.P. Goyal, Macmillan Business Books.

Data Science for Managerial Decisions (MB 511)

EMBA Program
MB-511

Online Resources

Software Resources

Data Science for Managerial Decisions (MB 511)

EMBA Program
MB-511

Program Overview

- Introduction to Data Science
- Information Technology An Overview
- Applications of Data Science in various fields
- MIS and Control Systems
- Data Collection and Data Pre-Processing
- Building Information Systems
- Support Systems for Management Decisions

Data Collection and Data Pre-Processing

EMBA Program
MB-511

- Introduction to Data Collection
- Methods of Data Collection in Management
- Designing Data Collection Instruments
- Sampling Techniques
- Data Collection Planning and Management
- Understanding Data Pre-Processing
- Data Cleaning Techniques
- Quality Assurance in Data Pre-Processing

Data Collection and Data Pre-Processing

Data Collection Planning and Management

EMBA Program
MB-511

Efficient data collection planning is crucial for ensuring that you gather high-quality data while optimizing resources such as time, money, and manpower. Here's a step-by-step methodology to efficiently plan data collection:

- Define Objectives and Research Questions: Clearly articulate the purpose of your data collection effort. What specific information are you trying to obtain? What research questions are you seeking to answer? Ensure that your objectives are well-defined and align with your overall research goals.

- Identify Data Needs: Determine the type of data required to address your research questions. This could include qualitative data (e.g., interviews, focus groups) or quantitative data (e.g., surveys, experiments). Consider whether existing data sources can fulfill your needs or if new data must be collected.

Data Collection and Data Pre-Processing

Data Collection Planning and Management

EMBA Program
MB-511

- Select Appropriate Data Collection Methods: Choose methods that are best suited to your research objectives, target population, and available resources. Common methods include surveys, interviews, observations, experiments, and secondary data analysis. Consider the advantages and limitations of each method in relation to your study.

- Design Data Collection Instruments: Develop tools such as questionnaires, interview guides, or observation protocols tailored to your research objectives. Ensure that these instruments are clear, concise, and unbiased to minimize measurement error. Pilot-test your instruments with a small sample to identify and address any issues before full-scale implementation.

Data Collection and Data Pre-Processing

Data Collection Planning and Management

- Select Appropriate Data Collection Methods: Choose methods that are best suited to your research objectives, target population, and available resources. Common methods include surveys, interviews, observations, experiments, and secondary data analysis. Consider the advantages and limitations of each method in relation to your study.

- Design Data Collection Instruments: Develop tools such as questionnaires, interview guides, or observation protocols tailored to your research objectives. Ensure that these instruments are clear, concise, and unbiased to minimize measurement error. Pilot-test your instruments with a small sample to identify and address any issues before full-scale implementation.

Data Collection and Data Pre-Processing

Data Collection Planning and Management

EMBA Program
MB-511

- Train Data Collectors: Provide training to individuals responsible for collecting data to ensure consistency and reliability. Train them on proper techniques, ethical considerations, and data handling procedures. Emphasize the importance of maintaining confidentiality and obtaining informed consent from participants.

- Implement Data Collection: Execute your data collection plan according to the established timeline and procedures. Monitor progress closely to identify and address any issues or deviations from the plan. Maintain regular communication with data collectors to provide guidance and support as needed.

Data Collection and Data Pre-Processing

Data Collection Planning and Management

EMBA Program
MB-511

- Monitor Data Quality: Implement quality control measures to ensure the accuracy, completeness, and validity of collected data. Conduct regular checks to identify and correct errors or inconsistencies. Utilize validation techniques, such as double data entry or inter-rater reliability tests, to enhance data quality.
- Manage Data Securely: Establish protocols for storing, managing, and protecting collected data to safeguard against unauthorized access or loss. Adhere to relevant privacy regulations and ethical guidelines governing data handling practices. Implement encryption, access controls, and backup procedures to mitigate security risks.

Data Collection and Data Pre-Processing

Data Collection Planning and Management

- Analyze and Interpret Data: Once data collection is complete, analyze the collected data using appropriate statistical or qualitative analysis techniques. Interpret the findings in relation to your research questions and objectives. Present results clearly and accurately, using visualizations or narrative summaries to convey key insights.
- Disseminate Findings: Share your findings with relevant stakeholders through reports, presentations, or publications. Tailor the format and content to the interests and needs of your audience. Consider opportunities for dissemination beyond academic circles, such as policy briefs, public forums, or media outreach.

Data Collection and Data Pre-Processing

Understanding Data Pre-Processing

Data preprocessing plays a crucial role in management across various domains, including business, healthcare, finance, and more. It involves transforming raw data into a clean, structured format that is suitable for analysis and decision-making.

- Data Quality Improvement: Preprocessing helps to enhance the quality of data by identifying and correcting errors, inconsistencies, and missing values. This ensures that the data used for analysis is accurate, reliable, and trustworthy, which is essential for making informed decisions.
- Data Integration: In many management scenarios, data comes from multiple sources and may be in different formats or structures. Preprocessing involves integrating diverse datasets into a unified format, facilitating comprehensive analysis and holistic decision-making. This integration can involve standardizing variables, resolving naming inconsistencies, and merging datasets.

Data Collection and Data Pre-Processing

Understanding Data Pre-Processing

Normalization and Standardization: Preprocessing techniques such as normalization and standardization are used to scale numerical data to a common range or distribution. This ensures that variables with different units or scales contribute equally to analysis, preventing bias in decision-making processes.

Feature Selection and Engineering: Preprocessing includes selecting relevant features (variables) and engineering new features that may better represent the underlying patterns in the data. By reducing dimensionality and focusing on the most informative features, management practitioners can improve the efficiency and accuracy of their analyses and decision models.

Data Reduction: In scenarios where datasets are large and complex, preprocessing techniques such as data reduction (e.g., sampling, aggregation, or dimensionality reduction) can be applied to streamline analysis and improve computational efficiency. This enables managers to focus on the most critical information without being overwhelmed by data volume.

Data Collection and Data Pre-Processing

Understanding Data Pre-Processing

Handling Missing Data: Preprocessing involves handling missing data through techniques such as imputation, where missing values are estimated based on available information. By addressing missing data effectively, managers can ensure that analyses are not biased and decision-making is based on the most complete information available.

Outlier Detection and Treatment: Outliers, or data points that deviate significantly from the rest of the dataset, can distort analysis results and decision-making processes. Preprocessing includes identifying and either removing or correcting outliers to prevent them from unduly influencing management decisions.

Data Collection and Data Pre-Processing

Data Pre-Processing in Decision Science

EMBA Program
MB-511

In machine learning (ML), data preprocessing is a critical step that involves transforming raw data into a format suitable for training machine learning models. Effective data preprocessing can significantly impact the performance and accuracy of ML models.

Data Cleaning:

- **Handling missing values:** This involves strategies such as imputation (replacing missing values with estimated values), deletion (removing rows or columns with missing values), or using algorithms that can handle missing values directly.
- **Dealing with outliers:** Outliers can skew the results of ML models. Techniques like trimming, winsorization, or capping can be used to handle outliers appropriately.

Data Transformation:

- **Feature scaling:** Scaling features to a similar range can prevent features with larger magnitudes from dominating the learning process. Common scaling techniques include min-max scaling and standardization (mean normalization).
- **Logarithmic transformation:** Transforming skewed data distributions using logarithmic transformations can help improve model performance, especially for features with long tails.
- **Box-Cox transformation:** This is another technique used to stabilize variance and make the data more Gaussian-like, which can benefit models that assume normality.

Data Collection and Data Pre-Processing

Data Pre-Processing in Decision Science

Feature Encoding:

- **One-Hot Encoding:** Converting categorical variables into a binary format, where each category becomes a binary feature column.
- **Label Encoding:** Assigning a unique numerical value to each category in ordinal variables.
- **Target Encoding:** Encoding categorical variables based on the target variable's mean or other statistics.
- **Embedding:** Representing categorical variables as dense vectors, often used in neural network-based models.

Data Collection and Data Pre-Processing

Data Pre-Processing in Decision Science

Feature Engineering:

- Creating new features: Combining or transforming existing features to create new ones that might be more informative for the model.
- Dimensionality reduction: Techniques like Principal Component Analysis (PCA) or Singular Value Decomposition (SVD) can reduce the dimensionality of the feature space while preserving most of the information.

Data Splitting:

- Splitting the dataset into training, validation, and test sets to assess model performance and prevent overfitting.
- Cross-validation: Repeatedly splitting the data into different training and validation sets to obtain a more robust estimate of model performance.

Data Collection and Data Pre-Processing

Data Pre-Processing in Decision Science

Handling Imbalanced Data:

- Techniques such as oversampling (e.g., SMOTE), undersampling, or using algorithms that are robust to class imbalance can address imbalanced datasets.

Normalization and Regularization:

- Normalizing inputs to the model, such as image pixel values or text embeddings, to improve convergence and stability during training.
- Applying regularization techniques like L1 or L2 regularization to prevent overfitting by penalizing large parameter values.

Data Augmentation:

- Generating synthetic data samples by applying transformations like rotation, flipping, or scaling to existing data, particularly in image or text data scenarios.

Data Collection and Data Pre-Processing

Data Cleaning Techniques

EMBA Program
MB-511

Handling Imbalanced Data:

- Techniques such as oversampling (e.g., SMOTE), undersampling, or using algorithms that are robust to class imbalance can address imbalanced datasets.

Normalization and Regularization:

- Normalizing inputs to the model, such as image pixel values or text embeddings, to improve convergence and stability during training.
- Applying regularization techniques like L1 or L2 regularization to prevent overfitting by penalizing large parameter values.

Data Augmentation:

- Generating synthetic data samples by applying transformations like rotation, flipping, or scaling to existing data, particularly in image or text data scenarios.

Data Management using Python

EMBA Program
MB-511

Team Activity

Quiz and Assignment 3

- Quiz 3 : April 06, 2024
- Assignment 3 : April 11, 2024

EMBA Program
MB-511

Have a question?

Feel Free to Reach out at

- +91-88846-52929 (WhatsApp)
- anant.awasthi@outlook.com (E-Mail)