

**BUA 682**  
**Data Pre-Processing for Business Analytics**  
**Graduate School of Business**  
**Maine Business School**  
**University of Maine**  
**Term: TBD**

---

**Instructor:** Dr. Tim Lu  
**Office:** Law Building #602, USM Campus, Portland  
**Phone:** 207-780-4531  
**Office Hours:** TBD  
**Email:** yonggang.lu@maine.edu

---

**MISSION:**

The mission of the Maine Business School is to advance business knowledge, connect with our local and global communities, seek adaptive approaches for business processes, and foster a sense of professionalism, teamwork, and respect for ourselves and our constituents.

**COURSE DESCRIPTION:**

This course is designed to enhance student's understanding of data quality problems commonly encountered in business environments including but not limited to missing data, noisy data, and data biases. This course discusses mechanisms of these problems and their impact on data analysis and modeling results and presents how to solve these problems by using different data pre-processing techniques such as imputation, integration, normalization, and transformation. Students practice these techniques with business data sets using mainstream analytical software.

**COURSE OBJECTIVES:**

After completing this course, students are expected to identify significant data quality issues in business data collection, especially large-scale data, using different analytical methods, and develop a good understanding of underlying causes of the issues. Students are also expected to be able to apply appropriate analytical techniques to solve the issues and produce preprocessed data in a ready-to-analyze state.

**COURSE CONTENT OUTLINE:**

*Module 1 – The Nature of Data and Data Pre-processing Process*

- Data, Dataset, Database and Data Information
- Common Data Quality Issues
- Analytical Tools and Data Pre-processing
- Data Pre-processing as a Process

*Module 2 – Data Pre-processing Basic Models*

- Normalization and Transformation
- Sampling, Variability and Confidence
- Handling Non-numerical Data

*Module 3 – Dealing with Missing Values*

- Simple Techniques for Missing Data
- Maximum Likelihood Imputation Methods

*Module 4 – Dealing with Noisy Data*

- Noise Detection
- Noise Filtering

*Module 5* – Data Reduction  
The Curse of Dimensionality  
Feature Selection