BELLEVUE UNIVERSITY

Course Syllabus Part I
DSC 550 - Data Mining

3 Credit Hours

Course Description
Data often contains patterns and anomalies that only emerge at large scale. In this course, you will import, clean, manipulate, visualize, analyze, and model structured and unstructured data to extract this information. Model building topics covered include text sentiment analysis, regression, classification, and neural networks. Furthermore, you will learn how to perform feature dimensionality reduction and tune model hyperparameters. The knowledge learned in this course culminates in a term project.

Course Prerequisites
Recommended: DSC 540 Data Preparation

Course Objectives
Students who successfully complete this course should be able to:

1. Import, analyze, and visualize large data sets
2. Transform data and reduce data dimensionality in preparation for data mining models
3. Construct natural language processing models
4. Choose and implement appropriate algorithms for model building
5. Apply tuning methods to optimize model hyperparameters
6. Make conclusions based on analysis of data

Grading Scale

93 – 100% = A	87 – 89% = B+	77 – 79% = C+	67 – 69% = D+
90 – 92% = A-	83 – 86% = B	73 – 76% = C	63 – 66% = D
	80 – 82% = B-	70 – 72% = C-	60 – 62% = D-
			0 – 59% = F

Topics Outline

I.   Importing and Exploring Data
    A. Creating data frames
    B. Exploring, summarizing, and aggregating data
    C. Dealing with missing data
II. Data Visualization
    A. Creating and analyzing data visualizations
    B. Best data visualization practices
III. Text, Categorical, and Time/Date Data
    A. Sentiment analysis
    B. Preparing text for model building
    C. Working with categorical data
    D. Handling dates and times
IV. Regression models
    A. Data preparation
    B. Training, evaluating, and interpreting regression models
V.   Classification Models and Unsupervised Learning
    A. Data preparation
    B. Training, evaluating, and interpreting classification models
    C. Cluster models
VI. Dimensionality Reduction
    A. Feature extraction
    B. Feature selection
    C. Feature engineering
    D. Principal component analysis
    E. Data leakage
VII. Model Selection and Tuning
    A. Model selection
    B. Hyperparameter tuning
    C. Grid search
    D. Pipelines
VIII. Neural Networks
    A. Artificial neural networks
    B. Convolutional neural networks
    C. Image classification
IX. Implementation
    A. Creating and organizing a data mining project
    B. Performing all steps and analysis for a data mining project
    C. Making a conclusion

Effective Date of Syllabus: October 14, 2021