

# Teaching Plan for DMSM

Subject Title	Subject length in hours
Data Mining- Advanced Statistical Modeling (with holiday)	65

## Subject Description

This course provides students a series of both theoretical and practical understanding of comprehensive methods in the fields of data mining, machine learning and predictive analytics. It covers various advanced modeling procedures and applications such as classification, prediction, exploration, association rules and time series forecasting that are performed and illustrated with real data in analytics software. In addition, students are presented with hands-on, industry and business backgrounds in different use cases. The learning outcome is to make students acquire practical experience in conducting real-world data analytics projects.

Textbooks/manuals

Materials by Metro College

## Method of Evaluation

(e.g., graded homework, quizzes, projects, final examination, et cetera; the type, number, and % value of each)

Type	Number	% Value	Type	Number	% Value	Type	Number	% Value
Quiz	1	20	Exam	1	40			
Project Report & Presentation	1	40						

## Teaching Method

Method	Number of hours
<input checked="" type="checkbox"/> Lecture	60
<input type="checkbox"/> Computer Based Learning	
<input type="checkbox"/> Seminar	
<input type="checkbox"/> Supervised Practical (e.g. clinic)	
<input checked="" type="checkbox"/> Supervised Lab (e.g. computers)	5
<input type="checkbox"/> Distance Education	

## Location

- ☒ Classroom/Lab (College)  
☐ Classroom/Lab (Off-campus)  
☐ Practicum

Does this subject require a final examination or a formal evaluation?

☒ Yes ☐ No

If "Yes", indicate the passing mark: 60%

# Teaching Plan for DMSM

---

## Objectives

---

Theory (the key elements that a student is expected to know upon completion of subject)

Core Competency: Familiar with the data sampling types and methods.

1. Know reason and process for data sampling
2. Understand right sample size for different purposes
3. Know random and weighted sample
4. General knowledge of sampling methods and procedures in data analytics tools

Core Competency: Understand the procedures of using different modeling methods to deal with various use cases

1. Understand supervised and unsupervised learning.
2. Familiar with theoretical background and mechanism of data mining or machine learning
3. Know various data mining or machine learning procedures and methods
4. Know how to conduct model validation
5. Know how to implement and deliver the analytical results

Core Competency: Familiar with the data transformation, variable generation and selection

1. Understand feature selection.
2. Understand feature transformation.
3. Know specific methodologies for feature selection and transformation
4. Be aware of over fitting problems and solutions
5. Know co-linearity issues and solutions

Core Competency: Know how to conduct pre analysis for advanced modeling

1. Understand correlation study by univariate analysis.

# Teaching Plan for DMSM

---

2. Know different data exploration methods and graphical institution.
3. understand how to recognize and fix data error

## Objectives

---

Skill ( the key behaviors that a student is expected to be able to perform upon completion of the subject)

Core Competency: Able to use analytics tools to conduct feature engineering and variables optimization

1. Able to create new features based on statistical sense
2. Able to select key drivers based on various methods using analytics tools
3. Able to apply some procedures on attributes such as standardization, reduction, binning and clustering, etc.
4. Able to read and explain statistical output from software
5. Able to use SQL to finish features engineering and transformations by different data manipulations

Core Competency: Able to carry out univariate & bivariate analysis, correlation analysis using different analytics software

1. Able to obtain descriptive statistics
2. Able to conduct data analysis for general patterns prior to modeling
3. Able to filter explanatory variables
4. Able to handle missing values and data imputation with different methods
5. Able to find and deal with outliers

Core Competency: Able to draw random and weighted samples for data analysis based on the specific requests

1. Able to draw different samples with different methods.
2. Able to conduct power study to determine sample size
3. Able to adjust models based on biased sample

# Teaching Plan for DMASM

---

Core Competency Able to apply different data mining, machine learning and predictive modeling methods to solve various practical projects

1. Able to find run different procedures of data mining and machine learning in Python.
2. Able to explain software output
3. Able to validate models and provide meaningful results
4. Able to convert quantitative result into practical presentation
5. Able to make recommendation for model implementation

# Teaching Plan for DMSM

## Subject/Module Outline

Date	Contact Hours	Module Topic-Sub Topics
Day 1	<b><u>5.0L</u></b>          (5.0)	<b>Introduction to Predictive Analytics</b>  1) Understand predictive models, supervised and unsupervised learning, data mining and machining concepts and use cases [App F-K-15(1)]  2) Learn the process for conducting predictive analytics step by step. [App F-K-15(1,2)]  3) Learn the different types of predictive models and business background. [App F-K-11(1,2)] [App F-S-12(1,2)]  <b>Introducing Data Sampling, Target, Bivariate Analysis, Interaction and Transformation</b>  1) Understand and conduct data sampling. Understand target rate. Perform random and weighted sampling and testing power. [App F-K-14(1,2,3,4)] [App F-S-13(1,2,3)] Classroom exercises Homework exercises
Day 2	<b><u>5.0L+0.5T</u></b>          (5.00)	<b>Introducing Data Sampling, Target, Bivariate Analysis, Interaction and Transformation</b>  1) Understand and conduct data sampling. Understand target rate. Perform random and weighted sampling and testing power. [App F-K-14(1,2,3,4)] [App F-S-13(1,2,3)]  2) Learn to use software to conduct bivariate analysis, interaction and transformation. [App F-K-11(2)] [App F-S-12(1,2,3)]  3) Learn feature engineering and selection methods. [App F-S-11(1,3)] [App F-S-11(1, 3, 4)] Classroom exercises Tutorial Project kick off and project tasks for the first day Homework exercises
Day 3	<b><u>5.0L+0.5T</u></b>	<b>Introducing Data Exploration, Performance Checking and Univariate Analysis.</b> 1) Understand data exploration, performance checking and univariate analysis.

Date	Contact Hours	Module Topic-Sub Topics
		<p>Conduct analysis using software. Read and organize outputs [App F-K-11(1,2,3)] [App F-S-12(1,2,3)]</p> <p>2) Learn various performance measures. Obtain these measures using software. Understand statistics using tables and visualization. [App F-K-13(1,2)] [App F-S-12(2,3)]</p> <p>3) Learn data clean, missing values, imputation and variable generation [App F-K-11(1,2,3)] [App F-K-13(3)] [ App F-S-12(4,5)]</p> <p><b>Introducing Linear Regression and Application</b></p> <p>1) Understand theory, procedure and related statistics. [App F-K-15(2,3)] [App F-K-14(1,2)]</p> <p>Classroom exercises Project reporting and new project task orientation Tutorial Homework exercises</p>
Day 4	<b>5.0L+0.5T</b>	<p>2) Learn to use software to perform linear regression. Understand output and interpretation. Diagnose model by statistics and plots. Learn to detect outliers[App F-K-15(3,4,5)] [App F-K-11(1,4,5)] [App F-S-14(1,2)] [App F-S-12(1,2,3)]</p> <p>3) Apply linear regression model to use cases. Feature selection and cross validation [App F-K-15(4,5)] [App F-K-11(4,5)] [App F-S-14(3,4,5)]</p> <p>Classroom exercises Tutorial Project reporting and new project task orientation Homework exercises</p>
Day 5	<b>5.0L+0.5T</b>	<p><b>Introducing Logistic Regression and Application</b></p> <p>1) Understand theory, procedure and model usage. Learn difference between logistic and linear regression. Show practical cases [App F-K-15(2,3)] [App F-K-14(1,2)]</p> <p>1) Understand theory, procedure and model usage. Learn difference between logistic and linear regression. Show practical cases</p>

Date	Contact Hours	Module Topic-Sub Topics
		[App F-K-15(2,3)] [App F-K-14(1,2)]
		2) Learn to use software to perform logistic regression. Understand output and interpretation. Introduce association measure [App F-K-15(3,4,5)] [App F-K-11(1,4,5)] [App F-S-14(1,2)] [App F-S-12(1,2,3)]  Project reporting and new project task orientation  Quiz  Tutorial  Homework exercises
	(5.0)	
Day 6	<b><u>5.0L+0.5T</u></b>	<b>Introducing Logistic Regression and Application (<i>continued</i>)</b>  3) Apply logistic regression model to use cases Feature selection and cross validation by deciles chart and gain chart [App F-K-15(5)] [App F-K-11(4,5)] [App F-S-14(3,4,5)] [App F-S-13(3)] [App F-S-11(4, 5)] [App F-S-12(1, 2, 3)]  <b>Decision Tree and Application</b>  1) Understand decision tree and regression tree. Understand train, validation and test samples. Familiar with software interface to conduct tree analysis. [App F-K-14(2, 3)] [App F-K-15(1,2,3,4,5)] 2) Learn statistics in tree. Understand the rule for cutting and pruning tree, cutting threshold, over fitting problem, nodes, leafs and final rules of tree model [App F-K-15(1,2,3)]  Project reporting and new project task orientation  Tutorial  Homework exercises
	(5.0)	
Day 7	<b><u>5.0L+0.5T</u></b>	3) Learn to use software to build different types of tree by examples. [App F-S-14(1,2,3)]  4) Software to validate tree. Understand and obtain tree performance statistics and charts. Score new data using tree model [App F-K-15(4,5)] [App F-S-14(4,5)]

## Teaching Plan for DMSM

Date	Contact Hours	Module Topic-Sub Topics
		<b>Developing Time Series Model And Application</b> 1) Understand time series forecasting model. Introduce several examples and software procedures [App F-K-14(2, 3)] [App F-K-15(1,2,3,4,5)] Classroom exercises  Project reporting and new project task orientation  Tutorial  Homework exercises
	(5.0)	
Day 8	<b><u>5.0L+0.5T</u></b>	2) Learn some basic concepts for building time series model – stationarity, white noise, random walk, data difference, auto-correlation, partial auto-correlation and data visualization [App F-K-13(1,2)] [App F-K-15(2)] 3) Introduce auto regression, moving average regression model and application of procedures in different software [App F-K-15(2, 3)] [App F-K-11(1,2,3,4)] 4) Understand validation and performance tracking statistics and plots for the resulting time series forecasting model [App F-K-15(3)] [App F-K-11(1,2,3,4)]  5) Build time series forecasting model using SAS procedures and GUI forecasting system. [App F-S-11(4)] [App F-S-12(1,2,3,4,5)] [App F-S-14(3)]  Classroom exercises  Project reporting and new project task orientation Tutorial Homework exercises
	(5.0)	
Day 9		<b>Holiday</b>
	<b><u>5.0L+0.5T</u></b>	



[illegible]

## Teaching Plan for DMASM

---

Date	Contact Hours	Module Topic-Sub Topics
		Review for examination, Questions and Answers Examination

Notes: the hours in [ ] are the time with instructor's supervision and the hours in ( ) are the minimum time a student should spend on review, homework, project and pre-reading new contents after class.  
Legends in the course outline: **L**- Lecture; **T**- Tutorial