Subject Description Form

Subject Code	AMA4680
Subject Title	Statistical Machine Learning
Credit Value	3
Level	4
Pre-requisite/	Applied Linear Models for Finance Analytics (AMA2602) or Applied Linear Models (AMA3602) or Statistics for Data Science (AMA3631) or equivalent **and** Mathematical Methods for Data Science (AMA3001/AMA3701) or Further Mathematical Methods for Finance (AMA3723) or Further Mathematical Methods (AMA3724) or equivalent **and** Programming Fundamentals and Applications (COMP1012) or Principles of Programming (AMA2222/AMA2222A) or equivalent
Exclusion	Statistical Machine Learning (AMA3800)
Objectives	To provide a basic introduction to machine learning. To present fundamental concepts and algorithms for selected topics of machine learning, to provide the students with the necessary background for the application of machine learning to real problems, and to provide a starting point for students who are interested in pursuing research in machine learning or related fields.
Intended Learning Outcomes	Upon completion of the subject, students will be able to: a) Demonstrate mastery of the principles of machine learning b) Develop quantitative skills of machine learning and interpret the outcomes of machine learning algorithms. c) Identify, define, and formulate problems of machine learning in real applications and generate workable solutions to problems.
Subject Synopsis/ Indicative Syllabus	<u>Supervised Learning, Regression and Model Selection:</u> Generalized linear models; ridge regression, model selection, dimension reduction, principal component analysis, lasso; reproducing kernels, kernel ridge regression, regularization; stochastic gradient descent, online learning. <u>Supervised Learning, Classification:</u> Naive Bayes; decision trees; *k*-nearest-neighbor classifier; logistic regression; support vector machines, kernelized support vector machines; score-based classifiers, the receiver operating characteristic curve, AUC scores, imbalanced data. <u>Unsupervised Learning, Clustering:</u>

<table>
<tr><td></td><td>K-means; agglomerative hierarchical clustering.

<u>Other Selected Topics:</u>
Cross validation; artificial neural networks, back-propagation, introduction to deep learning algorithms; introduction to software packages for machine learning; random forests.</td></tr>
<tr><td>Teaching/Learning Methodology</td><td>The subject will mainly be delivered through lectures and tutorials in computer lab. The theoretical background and the real applications of learning algorithms are both emphasized.</td></tr>
<tr><td>Assessment Methods in Alignment with Intended Learning Outcomes</td><td>
<table>
<tr><td rowspan="2">Specific assessment methods/tasks</td><td rowspan="2">% weighting</td><td colspan="3">Intended subject learning outcomes to be assessed (Please tick as appropriate)</td></tr>
<tr><td>a</td><td>b</td><td>c</td></tr>
<tr><td>1. Assignments/Projects</td><td>15%</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>2. Midterm Test</td><td>25%</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>3. Final Exam</td><td>60%</td><td>✓</td><td>✓</td><td>✓</td></tr>
<tr><td>Total</td><td>100%</td><td colspan="3"></td></tr>
</table>

Explanation of the appropriateness of the assessment methods in assessing the intended learning outcomes:
This subject focuses on the machine learning algorithms. Some of the algorithms concern important mathematical and statistical background. So it is important and appropriate to set the assignments and a midterm test for continuous assessment, and a final exam.
This subject emphasizes the practice of statistical machine learning. The projects are appropriate for assessing the related intended learning outcomes in the continuous assessment.</td></tr>
<tr><td>Student Study Effort Required</td><td>
<table>
<tr><td>Class contact:</td><td></td></tr>
<tr><td>Lecture</td><td>26 Hrs.</td></tr>
<tr><td>Tutorial</td><td>13 Hrs.</td></tr>
<tr><td>Other student study effort:</td><td></td></tr>
<tr><td>Assignments/Projects</td><td>58 Hrs.</td></tr>
<tr><td>Self-study</td><td>30 Hrs.</td></tr>
<tr><td>Total student study effort</td><td>127 Hrs.</td></tr>
</table>
</td></tr>
<tr><td>Reading List and References</td><td><u>Textbooks</u>:
<table>
<tr><td>Han, J., Kamber, M., and Pei, J.</td><td>Data Mining: Concepts and Techniques, 3rd Edition.</td><td>Morgan Kaufmann, 2011</td></tr>
</table>
</td></tr>
</table>

	James, G., Witten, D., Hastie, T., and Tibshirani, R.	An Introduction to Statistical Learning	Springer 2013
	<u>References:</u>		
	Tan, P.N., Steinbach, M., and Kummar, V.	Introduction to Data Mining	Pearson 2006
	Hastie, T., Tibshirani, R, and Friedman, J.	The Elements of Statistical Learning	Springer 2009
	Kelleher, J.D., Namee M.B., D'Arcy, A.	Fundamentals of Machine Learning for Predictive Data Analysis	The MIT Press 2015
	Steinwart, I., Christmann A.	Support Vector Machines	Springer 2008
	Goodfellow I., Bengio Y., Courville A.	Deep Learning	The MIT Press 2016