Master of Data Science

Apply now for entry in September 2020

- Machine Learning
- Data Analytics
- Advanced Statistical Modelling
- Social Networks
- Cloud Computing
The amount and complexity of digital data have grown exponentially over the years. In answer to the desperate call for experts in processing such data, Department of Statistics and Actuarial Science, The University of Hong Kong, has recently launched a new Taught Master Programme, Master of Data Science (MDASC), jointly with the Department of Computer Science, to teach students how to analyze data and formulate data-driven strategies.

Programme Information
Master of Data Science (MDASC) is a taught master programme jointly offered by Department of Statistics and Actuarial Science and Department of Computer Science. Its interdisciplinarity promotes the applications of computer technology, operational research, statistical modelling, and simulation to decision-making and problem-solving in all organizations and enterprises within the private and public sectors.

The curriculum of the MDASC programme adopts a well-balanced and comprehensive pedagogy of both statistical and computational concepts and methodologies, underpinning applications that are not limited to business or a single field alone.

Target Students
It is a programme ideal for:
1. those whose interest in high-level analytical skills straddles the disciplinary divide between statistics and computational analytics; and
2. those who wish to pursue further study in the field of data science after studying science, social sciences, engineering, medical sciences, information systems, computing and data analytics in their undergraduate studies.

Programme Highlights
The core courses of the proposed MDASC programme mainly focus on both predictive and prescriptive concepts and methodologies with an effort to equip students with a solid foundation on both statistical and computational analyses, e.g.

- Computational intelligence
- Data science technology
- Machine learning
- Statistical models
- Data mining techniques
- Financial data analysis
- Marketing analytics
- Quantitative risk management and finance
- Network security
- Multimedia technologies
- Smart phone apps development

Course Highlights
The electives cover a broad range of contemporary topics and provide students with solid training in diverse and applied techniques used in data science, including but not limited to

- Cluster & cloud computing
- Data mining techniques
- Financial data analysis
- Marketing analytics
- Quantitative risk management and finance
- Network security
- Multimedia technologies
- Smart phone apps development

Suoxinda Scholarship in Data Science
Two scholarships recipients, each receiving HK$20,000, would be selected from the students entering the Master of Data Science programme on the basis of academic merit and admission interview performance.

Compulsory Courses (36 credits)

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMP7404</td>
<td>Computational intelligence and machine learning</td>
<td>6 credits</td>
</tr>
<tr>
<td>DASC7011</td>
<td>Statistical inference for data science</td>
<td>6 credits</td>
</tr>
<tr>
<td>DASC7104</td>
<td>Advanced database systems</td>
<td>6 credits</td>
</tr>
<tr>
<td>DASC7806</td>
<td>Deep learning</td>
<td>6 credits</td>
</tr>
<tr>
<td>STAT7102</td>
<td>Advanced statistical modeling</td>
<td>6 credits</td>
</tr>
<tr>
<td>STAT8003</td>
<td>Time series forecasting</td>
<td>6 credits</td>
</tr>
</tbody>
</table>

Disciplinary Electives (24 credits)*

<table>
<thead>
<tr>
<th>List A</th>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COMP7105</td>
<td>Advanced topics in data science</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>COMP7305</td>
<td>Cluster and cloud computing</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>COMP7503</td>
<td>Multimedia technologies</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>COMP7506</td>
<td>Smart phone apps development</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>COMP7507</td>
<td>Visualization and visual analytics</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>COMP7906</td>
<td>Introduction to cyber security</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>ICOM6044</td>
<td>Data science for business</td>
<td>6 credits</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>List B</th>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MATH4801</td>
<td>Topics in applied discrete mathematics</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>MATH4803</td>
<td>Topics in mathematical programming and optimization</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>STAT6003</td>
<td>Financial data analysis</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>STAT6015</td>
<td>Advanced quantitative risk management and finance</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>STAT6016</td>
<td>Spatial data analysis</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>STAT7008</td>
<td>Programming for data science</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>STAT8017</td>
<td>Data mining techniques</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>STAT8019</td>
<td>Marketing analytics</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>STAT8006</td>
<td>Statistical methods for network data</td>
<td>3 credits</td>
</tr>
<tr>
<td></td>
<td>STAT8307</td>
<td>Natural language processing and text analytics</td>
<td>3 credits</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>DASC7600</td>
<td>Data science project</td>
<td>12 credits</td>
</tr>
</tbody>
</table>

*Students who have completed the same courses in their previous studies in HKU, e.g. Master of Statistics or Master of Science in Computer Science may, on production of relevant transcripts, be permitted to select up to 24 credits of disciplinary electives from either List A or List B above if they are not able to find any suitable options from either of the lists of disciplinary electives.

Programme Curriculum
Commencing in September, the curriculum is composed of 72 credits of courses. Courses with 6 credits are offered in the first and second semesters while courses with 3 credits are normally offered in the summer semester. If a student selects a course whose contents are similar to a course for which he/she has taken in his/her previous study, the Department may not approve the selection in question. The curriculum is the same for both full-time and part-time study modes.

Suoxinda Scholarship in Data Science
Two scholarships recipients, each receiving HK$20,000, would be selected from the students entering the Master of Data Science programme on the basis of academic merit and admission interview performance.

Capstone Requirement (12 credits)

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>DASC7600</td>
<td>Data science project</td>
<td>12 credits</td>
</tr>
</tbody>
</table>

“Working with numbers is in great demand in the years to come, evidenced in the impressive growth outlook numbers that help bolster the rankings of jobs like Data Scientist and Statistician.”

Excerpts from the 2019 Jobs Related Report by CareerCast.com

“Structured and robust educational foundation in statistical and computational analyses, e.g.
- Computational intelligence
- Data science technology
- Machine learning
- Statistical models
- Data mining techniques
- Financial data analysis
- Marketing analytics
- Quantitative risk management and finance
- Network security
- Multimedia technologies
- Smart phone apps development

Disciplinary Electives (24 credits)*

<table>
<thead>
<tr>
<th>List A</th>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>COMP7105</td>
<td>Advanced topics in data science</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>COMP7305</td>
<td>Cluster and cloud computing</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>COMP7503</td>
<td>Multimedia technologies</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>COMP7506</td>
<td>Smart phone apps development</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>COMP7507</td>
<td>Visualization and visual analytics</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>COMP7906</td>
<td>Introduction to cyber security</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>ICOM6044</td>
<td>Data science for business</td>
<td>6 credits</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>List B</th>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MATH4801</td>
<td>Topics in applied discrete mathematics</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>MATH4803</td>
<td>Topics in mathematical programming and optimization</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>STAT6003</td>
<td>Financial data analysis</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>STAT6015</td>
<td>Advanced quantitative risk management and finance</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>STAT6016</td>
<td>Spatial data analysis</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>STAT7008</td>
<td>Programming for data science</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>STAT8017</td>
<td>Data mining techniques</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>STAT8019</td>
<td>Marketing analytics</td>
<td>6 credits</td>
</tr>
<tr>
<td></td>
<td>STAT8006</td>
<td>Statistical methods for network data</td>
<td>3 credits</td>
</tr>
<tr>
<td></td>
<td>STAT8307</td>
<td>Natural language processing and text analytics</td>
<td>3 credits</td>
</tr>
</tbody>
</table>

*Students who have completed the same courses in their previous studies in HKU, e.g. Master of Statistics or Master of Science in Computer Science may, on production of relevant transcripts, be permitted to select up to 24 credits of disciplinary electives from either List A or List B above if they are not able to find any suitable options from either of the lists of disciplinary electives.

Capstone Requirement (12 credits)

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>DASC7600</td>
<td>Data science project</td>
<td>12 credits</td>
</tr>
</tbody>
</table>

“Working with numbers is in great demand in the years to come, evidenced in the impressive growth outlook numbers that help bolster the rankings of jobs like Data Scientist and Statistician.”

Excerpts from the 2019 Jobs Related Report by CareerCast.com
The amount and complexity of digital data have grown exponentially over the years. In answer to the desperate call for experts in processing such data, Department of Statistics and Actuarial Science, The University of Hong Kong, has recently launched a new Taught Master Programme, Master of Data Science (MDASC), jointly with the Department of Computer Science, to teach students how to analyze data and formulate data-driven strategies.

Programme Information
Master of Data Science (MDASC) is a taught master programme jointly offered by Department of Statistics and Actuarial Science (host) and Department of Computer Science.

Its interdisciplinarity promotes the applications of computer technology, operational research, statistical modelling, and simulation to decision-making and problem-solving in all organizations and enterprises within the private and public sectors.

The curriculum of the MDASC programme adopts a well-balanced and comprehensive pedagogy of both statistical and computational concepts and methodologies, underpinning applications that are not limited to business or a single field alone.

Target Students
It is a programme ideal for:
1. those whose interest in high-level analytical skills straddles the disciplinary divide between statistics and computational analytics;
2. those who wish to pursue further study in the field of data science after studying science, social sciences, engineering, medical sciences, information systems, computing and data analytics in their undergraduate studies.

Reimbursable Course(s) by Continuing Education Fund (CEF)
The following courses have been included in the list of reimbursable courses for CEF purposes:

- STAT097 Data mining techniques
- STAT109 Marketing analytics

The mother programme (Master of Data Science) of these courses is recognized under the Qualification Framework (QF) Level 6.

Programme Highlights
- Joint programme offered by Department of Statistics and Actuarial Science and Department of Computer Science
- Interdisciplinary and comprehensive curriculum
- Solid foundation in statistical and computational analyses
- Students can select electives from Computer Science, Mathematics and Statistics
- Electives cover a broad range of contemporary topics
- Hands-on applications of methodologies with powerful software
- Capstone project with real-life scenario

Course Highlights
The core courses of the proposed MDASC programme mainly focus on both predictive and prescriptive concepts and methodologies with an effort to equip students with a solid foundation in statistical and computational analyses, e.g.
- Computational intelligence
- Data science technology
- Machine learning

The electives cover a broad range of contemporary topics and provide students with solid training in diverse and applied techniques used in data science, including but not limited to:
- Cluster & cloud computing
- Data mining techniques
- Financial data analysis
- Marketing analytics
- Quantitative risk management and finance
- Network security
- Multimedia technologies
- Smart phone apps development

Suoxinda Scholarship in Data Science
Two scholarship recipients, each receiving HK$20,000, would be selected from the students entering the Master of Data Science programme on the basis of academic merit and admission interview performance.

“Working with numbers is in great demand in the years to come, evidenced in the impressive growth outlook numbers that help bolster the rankings of jobs like Data Scientist and Statistician.”

Extracted from the 2019 Jobs Related Report by CareerCast.com

---

### Programme Curriculum

Commencing in September, the curriculum is composed of 72 credits of courses. Courses with 6 credits are offered in the first and second semesters while courses with 3 credits are normally offered in the summer semester. If a student selects a course whose contents are similar to a course for which he/she has taken in his/her previous study, the Department may not approve the selection in question. The curriculum is the same for both full-time and part-time study modes.

#### Compulsory Courses (36 credits)

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMP7404</td>
<td>Computational intelligence and machine learning</td>
<td>6 credits</td>
</tr>
<tr>
<td>DASC7011</td>
<td>Statistical inference for data science</td>
<td>6 credits</td>
</tr>
<tr>
<td>DASC7104</td>
<td>Advanced database systems</td>
<td>6 credits</td>
</tr>
<tr>
<td>DASC7806</td>
<td>Deep learning</td>
<td>6 credits</td>
</tr>
<tr>
<td>STAT7102</td>
<td>Advanced statistical modelling</td>
<td>6 credits</td>
</tr>
<tr>
<td>STAT8003</td>
<td>Time series forecasting</td>
<td>6 credits</td>
</tr>
</tbody>
</table>

#### Disciplinary Electives (24 credits)*

with at least 12 credits from List A and 12 credits from List B

**List A**

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>COMP7105</td>
<td>Advanced topics in data science</td>
<td>6 credits</td>
</tr>
<tr>
<td>COMP7305</td>
<td>Cluster and cloud computing</td>
<td>6 credits</td>
</tr>
<tr>
<td>COMP7503</td>
<td>Multimedia technologies</td>
<td>6 credits</td>
</tr>
<tr>
<td>COMP7506</td>
<td>Smart phone apps development</td>
<td>6 credits</td>
</tr>
<tr>
<td>COMP7507</td>
<td>Visualization and visual analytics</td>
<td>6 credits</td>
</tr>
<tr>
<td>COMP7906</td>
<td>Introduction to cyber security</td>
<td>6 credits</td>
</tr>
<tr>
<td>ICOM6044</td>
<td>Data science for business</td>
<td>6 credits</td>
</tr>
</tbody>
</table>

**List B**

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>MATH6502</td>
<td>Topics in applied discrete mathematics</td>
<td>6 credits</td>
</tr>
<tr>
<td>MATH6503</td>
<td>Topics in mathematical programming and optimization</td>
<td>6 credits</td>
</tr>
<tr>
<td>STAT6013</td>
<td>Financial data analysis</td>
<td>6 credits</td>
</tr>
<tr>
<td>STAT6015</td>
<td>Advanced quantitative risk management and finance</td>
<td>6 credits</td>
</tr>
<tr>
<td>STAT6016</td>
<td>Spatial data analysis</td>
<td>6 credits</td>
</tr>
<tr>
<td>STAT7008</td>
<td>Programming for data science</td>
<td>6 credits</td>
</tr>
<tr>
<td>STAT8017</td>
<td>Data mining techniques</td>
<td>6 credits</td>
</tr>
<tr>
<td>STAT8019</td>
<td>Marketing analytics</td>
<td>6 credits</td>
</tr>
<tr>
<td>STAT8006</td>
<td>Statistical methods for network data</td>
<td>6 credits</td>
</tr>
<tr>
<td>STAT8307</td>
<td>Natural language processing and text analytics</td>
<td>5 credits</td>
</tr>
<tr>
<td>MATH8502</td>
<td>Topics in applied discrete mathematics</td>
<td>6 credits</td>
</tr>
</tbody>
</table>

*Students who have completed the same courses in their previous studies in HKU, e.g. Master of Statistics or Master of Science in Computer Science may, on production of relevant transcripts, be permitted to select up to 24 credits of disciplinary electives from either List A or List B above if they are not able to find any untaken options from either of the lists of disciplinary electives.

#### Capstone Requirement (12 credits)

<table>
<thead>
<tr>
<th>Course Code</th>
<th>Course Title</th>
<th>Credits</th>
</tr>
</thead>
<tbody>
<tr>
<td>DASC7600</td>
<td>Data science project</td>
<td>12 credits</td>
</tr>
</tbody>
</table>

---
Course Description

**Compulsory Courses**

**COMP7004** Computing intelligence and machine learning (6 credits)

This course will teach a broad set of principles and tools that will provide the mathematical and algorithmic framework for tackling problems using Artificial Intelligence (AI) and Machine Learning (ML). AI and ML are highly interdisciplinary fields with impact in different applications, such as, biology, robotics, languages, economics, and computer science. The course is aimed at engineering of making intelligent machines, especially intelligent computer programs, while ML refers to the changes in systems that perform tasks associated with AI.

**Pre-requisites:** ML, but knowledge of data structures and algorithms, probability, linear algebra, and programming would be an advantage.

**Assessment:** One 2-hour written examination; 50% coursework and 50% examination

**DASC7011** Statistical inference for data science (6 credits)

Computing power has revolutionized the theory and practice of statistical inference. Numerical, novel statistical inference procedures are becoming an integral part of data science. By focusing on the interplay between statistical inference and methodologies for data science, this course reviews the main concepts underlying classical statistical inference, studies computer-intensive methods for conducting statistical inference, and examines issues concerning the implementation of inference drawn upon modern learning technologies. Contents include classical frequentist and Bayesian inferences, computer-intensive methods such as the EM algorithm, the bootstrap and the Markov chain Monte Carlo, large-scale hypothesis testing, high-dimensional modeling, and post-model-selection inference.

**Assessment:** One 2-hour written examination; 40% coursework and 60% examination

**DASC7104** Advanced database systems (6 credits)

The course will study and explore the models and techniques in database systems, with a focus on the aspects of big data analytics, algorithms, and system design & organisation. It will also survey the recent development and progress in selected areas. Topics include: query optimisation, spatial-spatiotemporal data management, multimedia and time-series data management, information retrieval and XML, data mining.

**Assessment:** One 2-hour written examination; 50% coursework and 50% examination

**DASC7106** Deep learning (6 credits)

Machine learning is a fast-growing field in computer science and deep learning is the cutting edge technique that enables machines to learn from large-scale and complex data sets. This course will first introduce fundamental machine learning techniques and will then focus on artificial neural networks and how to train and optimize them to solve challenging problems using deep learning techniques. Topics include: auto-encoder, deep belief networks, convolutional neural networks, deep reinforcement learning and unsupervised feature learning. Popular deep learning software, such as Caffe, Torch and TensorFlow will also be introduced.

**Assessment:** One 2-hour written examination; 50% coursework and 50% examination

**STAT8003** Time series forecasting (6 credits)

A time series consists of a set of observations on a random variable taken over time. Such series arise naturally in econometrics, finance, environmental research and many other disciplines. In addition to statistical modelling, the course deals with the prediction of future behaviour of these time series. This course differs in emphasis from the typical time series course, investigates various representations for them and studies the relative merits of different forecasting procedures.

**Assessment:** One 3-hour written examination; 40% coursework and 60% examination

**Disciplinary Electives**

**COMP7015** Advanced topics in data science (6 credits)

This course will introduce selected advanced computational methods and apply them to problems in data analysis and relevant applications.

**Assessment:** One 2-hour written examination; 50% coursework and 50% examination

**COMP7025** Cluster and cloud computing (6 credits)

This course offers an overview of current cluster and cloud technologies, and discusses various issues in the design and implementation of cluster and cloud systems. Topics include cluster computing and virtualization techniques, such as Xen, KVM used in modern data centers. We will discuss three types of Cloud computing platforms, including SaaS, PaaS, and IaaS, by providing motivating examples from companies such as Amazon, Google, and Microsoft, and introducing Hadoop MapReduce and Spark programming paradigms for large-scale data analysis.

**Assessment:** One 2-hour written examination; 50% coursework and 50% examination

**COMP7055** Multimedia technologies (6 credits)

This course presents fundamental concepts and emerging technologies for multimedia computing. Students are expected to learn how to develop various kinds of media communication, presentation, and manipulation techniques. At the end of course, students should acquire proper skill set to utilise, integrate and synthesise different information and data from multimedia sources for building specific multimedia applications. Topics include media data acquisition methods and techniques, nature of perceptually encoded information, and manipulation of multimedia content organization and analysis; trending technologies for future multimedia computing.

**Assessment:** One 2-hour written examination; 50% coursework and 50% examination

**COMP7065** Smart phone apps development (6 credits)

Smart phones have become very popular in recent years. For iPhones alone, CEO Tim Cook announced that Apple has sold 30 billion iPhones by July 2015. In addition to iPhones, there are also Android phones, Symbian phones as well as Windows phones. Smart phones play an important role in mobile communication and applications. Smart phones are powerful as they support a wide range of applications called apps. Most of the time, smart phone users just purchase their favorite apps wirelessly from the vendors. There is a great potential applications (called apps). Most of the time, smart phone users just purchase their favorite apps wirelessly from the vendors. There is a great potential for software developers to reach worldwide. This course aims at introducing the design issues of smart phone apps. For example, the smart phone screen is usually much smaller than the computer monitor. We have to pay special attention to this aspect in order to develop attractive and successful apps. Other topics include: app development environments and programming techniques such as Java for Android phones, Objective-C and Swift for iPhones will be introduced to facilitate students to develop their own apps.

**Pre-requisites:** Students should have basic programming knowledge, e.g. C++ or Java.

**Assessment:** One 2-hour written examination; 50% coursework and 50% examination

**COMP7107** Visualization and visual analytics (6 credits)

This course introduces the basic principles and techniques in visualization and visual analytics, and their applications. Topics include computer vision, human visual perception, color visualization, visual data mining, and spatiotemporal data management, multimedia and time-series data mining.

**Assessment:** One 2-hour written examination; 50% coursework and 50% examination

**COMP7201** Introduction to cyber security (6 credits)

The aim of the course is to introduce different methods of protecting information and data in the cyber world, including the privacy issue. Topics include introduction to security, cyber attacks and threats; cryptographic algorithms and applications; network security and infrastructure.

**Assessment:** One 2-hour written examination; 50% coursework and 50% examination

**ICOM044** Data science for business (6 credits)

The emerging discipline of data science combines statistical methods with computer science to solve problems in applied area. In this case we focus on how data science can be used to solve business problems especially those in electronic commerce. By its very nature e-commerce is able to generate large amounts of data and data mining methods are quite helpful for managers in turning this data into knowledge which in turn can be used to make better decisions. These data sets and their accompanying quantitative methods have the potential to dramatically change decision making in many areas of business. For example, ideas like interactive marketing, customer relationship management, and database marketing are pushing companies to utilize the information they collect about their customers in order to make better marketing decisions.

This course focuses on how data science methods can be applied to solve managerial problems in marketing and electronic commerce. Our emphasis is on developing a core set of principles that embody data science: empirical reasoning, exploratory and visual analysis, and predictive modeling. We use these core principles to understand many methods used in data mining and machine learning. Our strategy in this course is to survey several popular techniques and understand how they map into these core principles. These techniques are illustrated with case studies. However, the emphasis is not on the software for implementing these techniques but on understanding the inputs and outputs of these techniques and how they are used to solve business problems.

**Pre-requisites:** Students should not be taking or have taken STAT6017 Data mining techniques or equivalent

**Assessment:** One 2-hour written examination; 65% coursework and 35% examination

**STAT6013** Financial data analysis (6 credits)

This course aims at introducing statistical methodologies in analyzing financial data. Financial applications and statistical methodologies are intertwined in all lectures. Contents include: recent advances in modern portfolio theory, Copula, market microstructure and high frequency data analysis, RefTech applications with various computational tools such as artificial neural networks, Kalman filters and blockchain data analysis.

**Assessment:** One 2-hour written examination; 40% coursework and 60% examination

**STAT6015** Advanced quantitative risk management and finance (6 credits)

This course covers statistical methodologies and models of importance to risk management and finance and links finance theory to market practice via statistical modeling and decision making. Emphasis will be put on empirical analyses to address the discrepancy between finance theory and market data. Contents include: Elementary Stochastic Calculus, Basic Monte Carlo and Quasi Monte Carlo Methods, Variance Reduction Techniques, Simulating the value of options and the value-at-risk for risk management, Review of univariate volatility model, multivariate volatility models, value-at-risk and expected shortfall, back testing and stress testing, Extreme value theory for risk management.

**Assessment:** One 2-hour written examination; 25% coursework and 75% examination

**MATH5002** Topics in applied discrete mathematics (6 credits)

This course aims to provide students with the opportunity to study some further topics in applied discrete mathematics. A selection of topics in discrete mathematics applied in combinatorics and optimization (such as algebraic coding theory, cryptography, discrete optimization, etc.) The selected topics may vary from year to year.

**Pre-requisites:** Knowledge in introductory discrete mathematics. Students may be asked to present appropriate evidence of having met the pre-requisites for enrolling in this course.

**Assessment:** One 2.5-hour written examination; 50% coursework and 50% examination

**MATH5003** Topics in mathematical programming and optimization (6 credits)

A study in greater depth of some special topics in mathematical programming or optimization. It is mainly intended for students in Operations Research or related subject areas. The course covers a selection of topics which may include convex, quadratic, geometric, stochastic programming, or discrete combinatorial optimization. The selected topics may vary from year to year.

**Pre-requisites:** Knowledge in introductory mathematical programming and optimization. Students may be asked to present appropriate evidence of having met the pre-requisites for enrolling in this course.

**Assessment:** 100% coursework

**STAT1015** Financial data analysis (6 credits)

This course aims at introducing statistical methodologies in analyzing financial data. Financial applications and statistical methodologies are intertwined in all lectures. Contents include: recent advances in modern portfolio theory, Copula, market microstructure and high frequency data analysis, RefTech applications with various computational tools such as artificial neural networks, Kalman filters and blockchain data analysis.

**Assessment:** One 2-hour written examination; 40% coursework and 60% examination

**STAT1015** Advanced quantitative risk management and finance (6 credits)

This course covers statistical methodologies and models of importance to risk management and finance and links finance theory to market practice via statistical modeling and decision making. Emphasis will be put on empirical analyses to address the discrepancy between finance theory and market data. Contents include: Elementary Stochastic Calculus, Basic Monte Carlo and Quasi Monte Carlo Methods, Variance Reduction Techniques, Simulating the value of options and the value-at-risk for risk management, Review of univariate volatility model, multivariate volatility models, value-at-risk and expected shortfall, back testing and stress testing, Extreme value theory for risk management.

**Assessment:** One 2-hour written examination; 25% coursework and 75% examination
Course Description

Compulsory Courses

Course Code: STAT8003
Title: Computer intelligence and machine learning (6 credits)

This course will teach a broad set of principles and tools that will provide the mathematical and algorithmic framework for tackling problems using Artificial Intelligence (AI) and Machine Learning (ML). AI and ML are highly interdisciplinary fields with impact in different applications, such as, biology, robotics, languages, economics, and computer science. AI is the science and engineering of making intelligent machines, especially intelligent computer programs, while ML refers to the changes in systems that perform tasks associated with AI.

Topics may include a subset of the following: problem solving by search, heuristic (informed) search, constraint satisfaction, games, knowledge-based agents, supervised learning, unsupervised learning, learning theory, reinforcement learning and adaptical control.

Pre-requisites: Nil, but knowledge of data structures and algorithms, probability, linear algebra, and programming would be an advantage.

Assessment: One 2-hour written examination; 50% coursework and 50% examination

Course Code: DASC7011
Title: Statistical inference for data science (6 credits)

Computing power has revolutionized the theory and practice of statistical inference. Technically, novel statistical inference procedures are becoming an integral part of data science. By focusing on the interplay between statistical inference and methodologies for data science, this course reviews the main concepts underlying classical statistical inference, studies computer-intensive methods for conducting statistical inference, and examines important issues concerning statistical inference drawn upon modern learning technologies. Contents include classical frequentist and Bayesian inference, computer-intensive methods such as the EM algorithm, the bootstrap, the Monte Carlo, large-scale hypothesis testing, high-dimensional modeling, and post-model-selection inference.

Assessment: One 2-hour written examination; 40% coursework and 60% examination

Course Code: DASC7104
Title: Advanced database systems (6 credits)

This course will study the design and implementation of database systems, with a focus on the areas of big data analytics, algorithms, and system design & operation. It will also survey the recent development and progress in selected areas. Topics include: query optimization, spatial-temporal data management, multimedia and time-series data management, information retrieval and XML, data mining.

Assessment: One 2-hour written examination; 50% coursework and 50% examination

Course Code: DASC7006
Title: Deep learning (6 credits)

Machine learning is a fast-growing field in computer science and deep learning is the cutting edge technology that enables machines to learn from large-scale and complex data sources. This course will first introduce fundamental machine learning techniques and will then focus on artificial neural networks and how to train and optimize them to solve challenging problems using deep learning, specific named entity recognition, natural language, convolutional neural networks, deep reinforcement learning and unsupervised feature learning. Popular deep learning software, such as Caffe, Torch and Theano will be introduced.

Assessment: One 2-hour written examination; 50% coursework and 50% examination

Course Code: STAT8013
Title: Time series forecasting (6 credits)

A time series consists of a set of observations on a random variable taken over time. Such series arise naturally in econometrics, finance, environmental research and many other disciplines. In addition to statistical modeling, the course deals with the prediction of future behaviour of these time series. This course distills the core ideas of time series, investigates various representations for them and studies the relative merits of different forecasting procedures.

Assessment: One 3-hour written examination; 40% coursework and 60% examination

Disciplinary Electives

Course Code: COMP7015
Title: Advanced topics in data science (6 credits)

This course will introduce selected advanced computational methods and apply them to problems in data analysis and relevant applications.

Assessment: One 2-hour written examination; 50% coursework and 50% examination

Course Code: COMP7005
Title: Cluster and cloud computing (6 credits)

This course offers an overview of current cluster and cloud technologies, and discusses various issues in the design and implementation of cluster and cloud systems. Topics include cluster infrastructure, scalability and virtualization techniques (e.g., Xen, KVM) used in modern data centers. We will discuss three types of Cloud computing platforms, including Saas, Paas, and Iaas, by providing motivating examples from companies such as Google, Amazon, and Microsoft; and introduce Hadoop MapReduce and Spark parallel programming paradigms for large-scale data analysis.

Assessment: One 2-hour written examination; 50% coursework and 50% examination

Course Code: COMP7050
Title: Multimedia technologies (6 credits)

This course presents fundamental concepts and emerging technologies for multimedia computing. Students are expected to learn how to develop various kinds of media communication, presentation, and manipulation techniques. At the end of course, students should acquire proper skill set to utilize, integrate and synthesize different information and data from multimedia sources for building specific multimedia applications. Topics include media data acquisition methods and techniques, nature of perceptually encoded information, and manipulation of multimedia content and organization and analysis; trending technologies for future multimedia computing.

Assessment: One 2-hour written examination; 50% coursework and 50% examination

Course Code: COMP7006
Title: Smart phone apps development (6 credits)

Smart phones have become very popular in recent years. For iPhones alone, CEO Tim Cook announced that Apple has sold the billionth iPhone in July 2016. Smart phones play an important role in mobile communication and application (called apps). Most of the time, smart phone users just purchase applications. Smart phones are powerful as they support a wide range of applications, such as email, calendar, games, social networking, office applications (including word processors and spreadsheets), photo editing, music, and video. From a business perspective, the iPhone has created a new wave of mobile computing and has led to a revolution in the mobile computing industry. In addition, there are also Android phones, Symbian phones as well as Windows phones.

Smart phones play an important role in mobile communication and applications. Smart phones are powerful as they support a wide range of applications (called apps). Most of the time, smart phone users just purchase their favorite apps wirelessly from the vendors. There is a great potential for software developers to reach worldwide markets.

This course aims at introducing the design issues of smart phone apps. For example, the smart phone screen is usually smaller than the computer monitor. We have to pay special attention to this aspect in order to develop attractive and useful apps. Other topics development environments and programming techniques such as Java for Android phones, Objective-C and Swift for iPhones will be introduced to facilitate students to develop their own apps.

Pre-requisites: Students should have basic programming knowledge, e.g., C++ or Java.

Assessment: One 2-hour written examination; 50% coursework and 50% examination

Course Code: STAT8017
Title: Data mining techniques or equivalent

This course focuses on how data science methods can be applied to solve business problems especially those in electronic commerce. By its very nature e-commerce is able to generate large amounts of data and data mining techniques are quite helpful for managers in turning this data into knowledge which in turn can be used to make better decisions. These data sets and their accompanying quantitative methods have the potential to dramatically change decision making in many areas of business. For example, ideas like interactive marketing, customer relationship management, and database marketing are pushing companies to utilize the information they collect about their customers in order to make better marketing decisions.

This course focuses on how data science methods can be applied to solve managerial problems in marketing and electronic commerce. Our emphasis is developing a core set of principles that embody data science: empirical reasoning, exploratory and visual analysis, and predictive modeling. We use these core principles to understand many methods used in data mining and machine learning. Our strategy in this course is to survey several popular techniques and understand how they map into these core principles. These techniques are illustrated with case studies. However, the emphasis is not on the software for implementing these techniques but on understanding the inputs and outputs of these techniques and how they are used to solve business problems.

Pre-requisites: Students should not be taking or have taken STAT8017 Data mining techniques or equivalent

Assessment: One 2-hour written examination; 65% coursework and 35% examination

Course Code: COMP7906
Title: Introduction to cyber security (6 credits)

The aim of the course is to introduce different methods of protecting information and data in the cyber world, including the privacy issue. Topics include introduction to security, cyber attacks and threats; cryptographic algorithms and applications; network security and infrastructure.

Assessment: One 2-hour written examination; 50% coursework and 50% examination

Course Code: ICOM0044
Title: Data science for business (6 credits)

The emerging discipline of data science combines statistical methods with computer science to solve problems in applied areas. In this we focus on how data science can be used to solve business problems especially those in electronic commerce. By its very nature e-commerce is able to generate large amounts of data and data mining techniques are quite helpful for managers in turning this data into knowledge which in turn can be used to make better decisions. These data sets and their accompanying quantitative methods have the potential to dramatically change decision making in many areas of business. For example, ideas like interactive marketing, customer relationship management, and database marketing are pushing companies to utilize the information they collect about their customers in order to make better marketing decisions.

This course focuses on how data science methods can be applied to solve managerial problems in marketing and electronic commerce. Our emphasis is developing a core set of principles that embody data science: empirical reasoning, exploratory and visual analysis, and predictive modeling. We use these core principles to understand many methods used in data mining and machine learning. Our strategy in this course is to survey several popular techniques and understand how they map into these core principles. These techniques are illustrated with case studies. However, the emphasis is not on the software for implementing these techniques but on understanding the inputs and outputs of these techniques and how they are used to solve business problems.

Pre-requisites: Nil, but knowledge of introductory discrete mathematics.

Assessment: One 2-hour written examination; 50% coursework and 50% examination

Course Code: STAT8013
Title: Financial data analysis (6 credits)

This course covers statistical methods and models of importance to risk management and Finance literacy and links finance theory to market practice via statistical modeling and decision making. Emphasis will be put on empirical analyses to address the discrepancy between finance theory and market data. Contents include: Elementary stochastic Calculus, Basic Monte Carlo and Quasi Monte Carlo Methods, Variance Reduction Techniques, Simulating the value of options and the value-at-risk for risk management, Review of univariate volatility models, multivariate volatility models, value-at-risk and expected shortfall, back testing and stress testing, Extreme value theory for risk management.

Assessment: One 2-hour written examination; 25% coursework and 75% examination

Course Code: STAT8013
Title: Advanced quantitative risk management and Finance (6 credits)

This course covers statistical methods and models of importance to risk management and Finance literacy and links finance theory to market practice via statistical modeling and decision making. Emphasis will be put on empirical analyses to address the discrepancy between finance theory and market data. Contents include: Elementary stochastic Calculus, Basic Monte Carlo and Quasi Monte Carlo Methods, Variance Reduction Techniques, Simulating the value of options and the value-at-risk for risk management, Review of univariate volatility models, multivariate volatility models, value-at-risk and expected shortfall, back testing and stress testing, Extreme value theory for risk management.

Assessment: One 2-hour written examination; 25% coursework and 75% examination
The Big Data Research Cluster (BDRC) of HKU Faculty of Science, launched in 2016, aims to provide HKU wide leadership and strategies in research, education, and technology transfer in the multidisciplinary field of big data, and thereby generate bigger impacts to HKU and the broader community.

"Unlike traditional data, solution to issues in Big Data will usually call for expertise in different disciplines including but not limited to the owner of the data, IT experts, statisticians and experts in the relevant field," said Professor W.K. Li, former Director of BDRC.

Programme Duration
The programme normally extends over 1.5 academic years for full-time study, and 2 academic years for part-time study mode. Fast tracking is allowed under exceptional circumstances.

Optional Summer Courses
Tutorials on statistical software (e.g. R and SAS) and computer programming (e.g. Python) will be held in August, 2020 for students who need to rejuvenate their skills.

Programme Evaluation
Assessment: One 2-hour written examination; 50% coursework and 50% examination

STAT8019 Marketing analytics (6 credits)
This course introduces the methodology of text analytics. Topics include natural language processing, word representation, machine learning systems. This course introduces the methodology of text analytics. Topics include natural language processing, word representation, text categorization and clustering, topic modelling and sentiment analysis. Students are required to possess basic understanding of Python language.

Pre-requisites: Students should not be taking or have taken ICOM6044 Data science for business or equivalent
Assessment: 100% coursework

STAT8017 Data mining techniques (6 credits)
The course aims at helping people to work smarter by revealing the underlying structure and relationships in large amounts of data. The course also will help students to be able to work with a variety of sources (e.g. excel, web-scraping, APIs and others), object-oriented programming concepts and scientific computation tools. Students will learn to create their own R packages or Python libraries.

Assessment: 100% coursework

STAT8017 Data mining techniques (6 credits)
The course aims at helping people to work smarter by revealing the underlying structure and relationships in large amounts of data. The course also will help students to be able to work with a variety of sources (e.g. excel, web-scraping, APIs and others), object-oriented programming concepts and scientific computation tools. Students will learn to create their own R packages or Python libraries.

Assessment: 100% coursework

STAT8018 Natural language processing and text analytics
The course introduces the methodology of text analytics. Topics include natural language processing, word representation, text categorization and clustering, topic modelling and sentiment analysis. Students are required to possess basic understanding of Python language.

Pre-requisites: Pass in STAT8017 Data mining techniques or equivalent
Assessment: 100% coursework

Capstone Requirement
A project will be required to carry out independent work on a major project under the supervision of individual staff member. A written report is required.

Assessment: 75% written report and 25% oral presentation

Testimonials
As a quantitative researcher, I have to apply a lot of knowledge in finance, risk management and statistics in my job. Therefore, after pursuing the FRM and CFA qualification, I started my MSc study in 2013 in order to strengthen my statistical knowledge in the areas of finance and risk management. I learned a lot of useful statistical techniques during my MSc study.

Since practice is the sole criterion for verifying the truth, the best way to practice the theories and techniques I learned from MSc is to implement them in my job. To achieve this goal, I used data analysis and programming knowledge as essential tools. Thus, I joined the MSc/CO programme this year. It is good to know HKU has launched a new master programme, Master of Data Science, which covers the latest statistical techniques like machine learning and big data analytics. The programme is also designed in a practical manner to allow students to solidify their knowledge through implementation.

I use the knowledge I gained from the course on computational statistics, and I would be happy to recommend the people who are interested in quantitative research to do the programme.

WONG Cheuk Lam, CFA, FRM
Head of Research, Temujin Capital Management

To succeed in data science in the commercial world, you must be excellent in the fields of Computer Science, Statistics and Business Intelligence. That was what I found for a long time in my career. As a computer engineering graduate, I started my career in Data Science and gradually found there was an obstacle to have advancement on my career. I realized that I need more solid skills and knowledge in Statistics to understand the algorithms and perfect my analyses.

I joined the MSc(CS) programme this year. It aims at discussing the common properties of real networks and the recent development of network models. Topics include common network measures, community detection in graphs, preferential attachment random network models, exponential random graph models, models based on random point processes and the hidden network discovery on a set of dependent random variables.

Assessment: 1-1.5 hour written examination; 50% coursework and 50% examination

STAT8016 Statistical methods for network data (3 credits)
The six degrees of separation theories that human interactions could be essentially represented in the form of a network. Examples of networks include router networks, the World Wide Web, social networks (e.g. Facebook or Twitter), genetic interaction networks and various collaboration networks (e.g. movie actor collaboration network and scientific paper collaboration network). Despite the diversity in the nature of sources, the networks exhibit some common properties. For example, both the spread of disease in a population and the spread of rumors in a social network are in sub-logarithmic time. This course aims at discussing the common properties of real networks and the recent development of statistical network models. Topics may include common network measures, community detection in graphs, preferential attachment random network models, exponential random graph models, models based on random point processes and the hidden network discovery on a set of dependent random variables.

Assessment: 1-hour written examination; 50% coursework and 50% examination

STAT8015 Data science for business or equivalent
Assessment: 75% written report and 25% oral presentation

STAT8017 Data mining techniques (6 credits)
This course introduces various statistical models and methodology used in market research. Special emphasis will be put on market analytics and statistical techniques for marketing decision making including market segmentation, market response models, consumer preference analysis and pricing models. Contents include market response models, statistical methods for segmentation, targeting and positioning, statistical methods for new product design.

Assessment: One 3-hour written examination; 40% coursework and 60% examination

Optional Summer Courses
Tutorials on statistical software (e.g. R and SAS) and computer programming (e.g. Python) will be held in August, 2020 for students who need to rejuvenate their skills.

Data Science helps us decipher the data and extract valuable information that can be used as a strategic part of critical decision-making. It is therefore no surprise that Data Scientists are of huge demand by employers all over the world. As an international premier centre for excellence, HKU always strives to provide students with a state-of-the-art education. The launch of Master of Data Science (MDAS) provides students with the chance to acquire the quantitative and analytical skills much needed to stay ahead of the game.

Cheung Ting Hin (MDAS Part-time Student)
Head of Research, Temujin Capital Management

Being a part-time MDAS student with no statistic nor programming background, it was definitely a challenging yet fruitful experience so far. After over a decade working in the sales and trading industry, I felt refreshing back to the campus and studying all the fancy formula and symbols again.

I would say the courses are much more demanding than I expected, but thanks to the summer preparation classes, it helped to recall my memories on some basic concepts.

Cheung Ting Hin (MDAS Part-time Student)
Director, RRMC Hong Kong
Data Science helps us decipher the data and extract valuable information that can be used as a strategic part of critical decision-making. It is therefore no surprise that Data Scientists are of huge demand by employers all over the world.

As an international premier centre par excellence, HKU always strives to provide students with a state-of-the-art education. The launch of Master of Data Science (MDaS) provides students with the chance to acquire the quantitative and analytical skills much needed to stay ahead of the game.

Testimonials
As a quantitative researcher, I have to apply a lot of knowledge in finance, risk management and statistics in my job. Therefore, after pursuing the FRM and CPA qualification, I started my MDaS study in 2013 in order to strengthen my statistical knowledge in the fields of finance and risk management. I learned a lot of useful statistical techniques during my MDaS study. Since practice is the sole criterion for verifying the truth, the best way to practice the theories and techniques I learned from MDaS is to implement them in my job. To achieve this goal, I worked with large computer and programming knowledge are essential. Thus, I joined the MDaS programme this year. It is good to know HKU has launched a new master programme, Master of Data Science, which covers the latest statistical techniques like machine learning and big data analysis. The programme is also designed in a practical manner to allow students to solidify their knowledge through implementation. I am excited for the opportunity and I am sure that I will be happy to recommend the programme to the people who are interested in quantitative research to do this programme.

WONG Cheuk Lam, CPA, FRM
Head of Research, Tianrui Capital Management

Big Data in HKU

The HKU Innovative Data Mining Application Award 2019, sponsored by Shenzhen Suxianda Industry Company Ltd, announced the winners in May 2019. The winning teams produced impressive projects which showcased their ability to apply data mining techniques to solve real-life challenges. The team winning first prize conducted “An Analysis of Fake News Detection” to examine and compare the performance of fake news classifiers built upon various natural language processing techniques to solve real-life challenges.

The team winning second prize conducted “An Analysis of Fake News Detection” to examine and compare the performance of fake news classifiers built upon various natural language processing techniques to solve real-life challenges. To succeed in data science in the commercial world, you must be excellent in the fields of Computer Science, Statistics and Business Intelligence. That was what I found at the beginning of my career. As a computer science graduate, I started my career in Data Science and gradually found there was an obstacle to have advanced on my career. I realized that I should gain more solid skills and knowledge in Statistics to understand the algorithms and perfect my analyses. I chose Information and Statistics programme from the University of Hong Kong, which was one of the most important decisions in my life. Data are always around us. The key is not data. Instead, it’s how you handle it and what you get from it. The MStat programme provided me with the practical matter to allow students to solidify their knowledge through implementation. It was the edge of both MDaS and the MStat programme and I would be happy to recommend the programme to the people who are interested in quantitative research to do this programme.

WONG Cheuk Lam, CPA, FRM
Head of Research, Tianrui Capital Management

Programme Duration
The programme normally extends over 1.5 academic years for full-time study, and 2.5 academic years for part-time study. Fast-track completion is possible for both study modes, subject to the approval of the University. Students could graduate in 1 year for full-time study mode, and 2 years for part-time study mode. All lectures are conducted in English at HKU.

Optional Summer Courses
Tutorials on statistical software (e.g. R and SAS) and computer programming (e.g. Python) will be held in August, 2020 for students who need to reinforce their skills.
Tuition Fees
The full tuition fees for the programme is HK$252,000 for the 2020 intake. The fee shall normally be payable in three instalments over 1.5 years for full-time study or in five instalments over 2.5 years for part-time study. In addition, students are required to pay Caution Money (HK$350), refundable on graduation subject to no claims being made, and Graduation Fee (HK$350).

Subject to approval

Admission Requirements
- Applicants shall hold a Bachelor’s degree with Honours or an equivalent qualification;
- Applicants should have taken at least one university or post-secondary certificate course in each of the following three subjects (calculus and algebra, computer programming and introductory statistics) or related areas.
- Applicants shall fulfil the University Entrance Requirements.

Application
Main Round: December 15, 2019
Clearing Round: 12 noon, January 31, 2020

Programme Details:
https://saasweb.hku.hk/programme/mdasc-index.php

Online Application:
https://aal.hku.hk/tpg/