



**ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH**  
**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**ĐỀ CƯƠNG MÔN HỌC**  
**DS102 – HỌC MÁY THỐNG KÊ**

**1. THÔNG TIN CHUNG (General information)**

Tên môn học (tiếng Việt):	<b>Học máy thống kê</b>
Tên môn học (tiếng Anh):	Statistical Machine Learning
Mã môn học:	<b>DS102</b>
Thuộc khối kiến thức:	Đại cương <input type="checkbox"/> ; Cơ sở nhóm ngành <input type="checkbox"/> ; Cơ sở ngành <input checked="" type="checkbox"/> ; Chuyên ngành <input type="checkbox"/> ; Tốt nghiệp <input type="checkbox"/>
Khoa, Bộ môn phụ trách:	Khoa Khoa học và Kỹ thuật Thông tin
Giảng viên biên soạn:	TS. Nguyễn Lưu Thuỳ Ngân (nganlt@uit.edu.vn) ThS. Nguyễn Văn Kiệt (kietnv@uit.edu.vn) ThS. Nguyễn Đức Vũ (vund@uit.edu.vn) CN. Lưu Thanh Sơn (sonlt@uit.edu.vn) CN. Huỳnh Văn Tín (tinhv@uit.edu.vn)
Số tín chỉ:	
Lý thuyết:	3 .....
Thực hành:	1 .....
Tự học:	6 .....
Môn học tiên quyết:	Không .....
Môn học trước:	Không .....

**2. MÔ TẢ MÔN HỌC (Course description)**

Môn học cung cấp cho sinh viên các kiến thức cơ bản về máy học như: định nghĩa về máy học, đánh giá mô hình, bài toán phân lớp, bài toán hồi quy, các mô hình học có giám sát và không giám sát. Đồng thời môn học cũng giới thiệu cho sinh viên các bài toán thực tế cần sử dụng máy học để giải quyết.

**3. MỤC TIÊU MÔN HỌC (Course goals)**

Sau khi hoàn thành môn học này, sinh viên có thể:

Ký hiệu	Mục tiêu môn học	Chuẩn đầu ra trong CTĐT
<i>G1</i>	Biết và hiểu được định nghĩa máy học, các bài toán áp dụng máy học và các kỹ thuật dùng trong máy học	<i>LO2, LO3, LO4</i>
<i>G2</i>	Áp dụng kiến thức về máy học vào giải quyết một bài toán máy học cụ thể trong thực tế.	<i>LO4, LO10</i>

#### 4. CHUẨN ĐẦU RA MÔN HỌC (Course learning outcomes)

CĐRMH	Mô tả CĐRMH (Mục tiêu cụ thể)	Mức độ giảng dạy
<i>G1.1</i>	Biết được các khái niệm căn bản: máy học, bài toán phân lớp, bài toán hồi quy.	<i>TU</i>
<i>G1.2</i>	Hiểu cách hoạt động của các thuật toán máy học, cách đánh giá một mô hình máy học.	<i>TU</i>
<i>G2.1</i>	Biết được một số bài toán về máy học trong thực tế.	<i>IT</i>
<i>G2.2</i>	Vận dụng các kỹ thuật máy học đã học vào giải quyết cho bài toán.	<i>IT</i>

#### 5. NỘI DUNG MÔN HỌC, KẾ HOẠCH GIẢNG DẠY (Course content, lesson plan)

##### a. Lý thuyết

Buổi học (30 tiết)	Nội dung	CĐRMH	Hoạt động dạy và học	Thành phần đánh giá
Buổi 1 (4 tiết)	<b>Chương 1: Tổng quan</b> 1.1. Định nghĩa máy học. 1.2. Quy trình thực hiện một dự án máy học. 1.3. Các dạng máy học: - Học có giám sát. - Học không giám sát. - Học tăng cường.	<i>G1.1</i>	<b>Dạy:</b> GV giải thích các định nghĩa. <b>Học ở lớp:</b> Sinh viên theo dõi bài giảng và trả lời câu hỏi của giảng viên. <b>Học ở nhà:</b> Đọc thêm Chương 1 trong sách.	<i>A1</i>
Buổi 2 (4 tiết)	<b>Chương 2: Triển khai một dự án máy học</b> 2.1. Định nghĩa tác vụ. 2.2. Bộ dữ liệu phục vụ cho bài toán. 2.3. Chọn mô hình.	<i>G1.1</i> <i>G1.2</i>	<b>Dạy:</b> GV giải thích các định nghĩa. <b>Học ở lớp:</b> Sinh viên theo dõi bài giảng và trả lời	<i>A1</i>

	2.4. Phân chia dữ liệu. 2.5. Đánh giá mô hình.		câu hỏi của giảng viên. <b>Học ở nhà:</b> Đọc thêm Chương 2 trong sách.	
Buổi 3 (4 tiết)	<b>Chương 3: Bài toán phân lớp</b> 3.1. Định nghĩa bài toán. 3.2. Các dạng bài toán phân lớp: - Phân lớp nhị phân. - Phân lớp đa nhãn. - Phân lớp đa lớp. 3.3. Các độ đo dùng cho bài toán phân lớp. - Precision. - Recall. - F1-score. 3.4. Phân tích lỗi. 3.5. Case study: <i>Phân loại ảnh X-quang phổi</i> .	<i>G1.1</i> <i>G1.2</i> <i>G2.1</i>	<b>Dạy:</b> GV giới thiệu các định nghĩa và các mô hình. <b>Học ở lớp:</b> Sinh viên theo dõi bài giảng và trả lời câu hỏi của giảng viên. <b>Học ở nhà:</b> Sinh viên đọc thêm Chương 3 trong sách.	<i>AI</i>
Buổi 4 (4 tiết)	<b>Chương 4: Hồi quy và huấn luyện mô hình.</b> 4.1. Định nghĩa bài toán. 4.2. Các mô hình hồi quy - Hồi quy tuyến tính. - Hồi quy đa biến. - Hồi quy Ridge. - Hồi quy logistic. 4.3. Độ đo dùng cho bài toán hồi quy.	<i>G1.1</i> <i>G1.2</i> <i>G2.1</i>	<b>Dạy:</b> GV giới thiệu các định nghĩa và các mô hình. <b>Học ở lớp:</b> Sinh viên theo dõi bài giảng và trả lời câu hỏi của giảng viên. <b>Học ở nhà:</b> Sinh viên đọc thêm Chương 4 trong sách.	<i>AI</i>
Buổi 5 (4 tiết)	<b>Chương 5: Gradient descent</b> 5.1. Giới thiệu về gradient descent. 5.2. Batch gradient descent (BGD) 5.3. Stochastic Gradient descent. 5.4. Mini-batch gradient descent (MGD)	<i>G1.1</i> <i>G1.2</i> <i>G2.1</i>	<b>Dạy:</b> GV giới thiệu các định nghĩa và các mô hình. <b>Học ở lớp:</b> Sinh viên theo dõi bài giảng và trả lời câu hỏi của giảng viên. <b>Học ở nhà:</b> Sinh viên đọc thêm Chương 4 trong sách.	<i>AI</i>
Buổi 6 + 7	<b>Chương 6: Một số thuật toán máy học dành cho phân</b>		<b>Dạy:</b> GV giới thiệu các định	

(8 tiết)	<b>lớp</b> - Naive Bayes - KNN - SVM - Decision Tree		nghĩa và các mô hình. <b>Học ở lớp:</b> Sinh viên theo dõi bài giảng và trả lời câu hỏi của giảng viên. <b>Học ở nhà:</b> Sinh viên đọc thêm Chương 5 và Chương 6 trong sách.	
Buổi 8 (4 tiết)	<b>Chương 7: Ensemble model</b> 7.1. Giới thiệu về ensemble model. 7.2. Các phương pháp ensemble. - Voting. - Random forest - Boosting.	<i>G1.1</i> <i>G1.2</i> <i>G2.1</i>	<b>Dạy:</b> GV giới thiệu các định nghĩa và các mô hình. <b>Học ở lớp:</b> Sinh viên theo dõi bài giảng và trả lời câu hỏi của giảng viên. <b>Học ở nhà:</b> Sinh viên đọc thêm Chương 7 trong sách.	<i>A1</i>
Buổi 9 (4 tiết)	<b>Chương 8: Giảm chiều dữ liệu.</b> 8.1. Sự cần thiết của giảm chiều. 8.2. Các chiến lược giảm chiều. 8.3. Thuật toán giảm chiều PCA. 8.4. Thuật toán giảm chiều LLE			
Buổi 10 (4 tiết)	<b>Chương 9: Giới thiệu sơ lược về Neural network.</b> 9.1. Các khái niệm cơ bản. - Units. - Layers. - ANN. 9.2. Mạng neural 2 lớp. 9.3. Huấn luyện mạng neural. 9.4. Case study.	<i>G1.1</i> <i>G1.2</i> <i>G2.1</i>	<b>Dạy:</b> GV giới thiệu các định nghĩa và các mô hình. <b>Học ở lớp:</b> Sinh viên theo dõi bài giảng và trả lời câu hỏi của giảng viên. <b>Học ở nhà:</b> Sinh viên đọc thêm Chương 10+11 trong sách.	<i>A1</i>
Buổi 11	<b>Chương 10: Bài toán phân</b>	<i>G1.1</i>	<b>Dạy:</b> Giảng viên	<i>A4</i>

(4 tiết)	<b>cụm.</b> 10.1. Bài toán phân cụm và học máy không giám sát. 10.2. Các thuật toán phân cụm: KMeans, DBScan. 10.3. Mô hình hỗn hợp Gauss.	<i>G1.2</i> <i>G2.1</i> <i>G2.2</i>	chấm đồ án cho SV tại lớp. <b>Học ở lớp:</b> Sinh viên báo cáo thuyết trình cho đồ án.	
----------	---	---	---	--

## b. Thực hành

Hình thức thực hành: 1

Buổi học (30 tiết)	Nội dung	ĐĐRMH	Hoạt động dạy và học	Thành phần đánh giá
Buổi 1 (5 tiết)	<b>Bài thực hành 1:</b> Máy học cơ bản Phần 1 - Làm quen với python và môi trường phát triển dành cho python (Google colab, Jupyter notebook) - Hiện thực Linear Regression và Ridge Regression bằng tay sử dụng các thư viện cơ bản như numpy.	<i>G2.1</i> <i>G2.2</i>	<b>Dạy:</b> GV hướng dẫn sinh viên tại lớp. <b>Học ở lớp:</b> Sinh viên thực hành theo hướng dẫn của giảng viên. <b>Học ở nhà:</b> Sinh viên hoàn thành bài thực hành theo yêu cầu.	<i>A3</i>
Buổi 2 (5 tiết)	<b>Bài thực hành 2:</b> Máy học cơ bản Phần 2 - Xây dựng mô hình Hồi quy Logistic bằng tay. - Hiện thực thuật toán Gradient Descent. - Dự đoán trên bộ dữ liệu Iris.	<i>G2.1</i> <i>G2.2</i>	<b>Dạy:</b> GV hướng dẫn sinh viên tại lớp. <b>Học ở lớp:</b> Sinh viên thực hành theo hướng dẫn của giảng viên. <b>Học ở nhà:</b> Sinh viên hoàn thành bài thực hành theo yêu cầu.	<i>A3</i>
Buổi 3 (5 tiết)	<b>Bài thực hành 3:</b> Dự án máy học cơ bản - Hồi quy tuyến tính - Sử dụng bộ dữ liệu House Price California cho bài toán dự đoán giá nhà. - Đọc và phân tích khám phá dữ liệu. - Xây dựng mô hình hồi quy tuyến tính cho bài toán. - Đánh giá mô hình bằng độ đo MSE	<i>G2.1</i> <i>G2.2</i>	<b>Dạy:</b> GV hướng dẫn sinh viên tại lớp. <b>Học ở lớp:</b> Sinh viên thực hành theo hướng dẫn của giảng viên. <b>Học ở nhà:</b> Sinh viên hoàn thành bài thực hành theo yêu cầu.	<i>A3</i>
Buổi 4 (5 tiết)	<b>Bài thực hành 4:</b> Bài toán phân lớp.	<i>G2.1</i> <i>G2.2</i>	<b>Dạy:</b> GV hướng dẫn sinh viên tại	<i>A3</i>

	<ul style="list-style-type: none"> <li>- Sử dụng bộ dữ liệu Iris và MNIST cho bài toán phân lớp.</li> <li>- Xây dựng các mô hình phân lớp cơ bản gồm: Logistic Regression, KNN, SVM.</li> <li>- Đánh giá hệ thống phân lớp bằng các độ đo.</li> <li>- Sử dụng các chiến lược tinh chỉnh tham số.</li> <li>- Phân tích lỗi.</li> </ul>		lớp. <b>Học ở lớp:</b> Sinh viên thực hành theo hướng dẫn của giảng viên. <b>Học ở nhà:</b> Sinh viên hoàn thành bài thực hành theo yêu cầu.	
Buổi 5 (5 tiết)	<b>Bài thực hành 5:</b> Bài toán phân lớp văn bản. <ul style="list-style-type: none"> <li>- Bộ dữ liệu: UIT-VSFC.</li> <li>- Các mô hình áp dụng: Naive Bayes, KNN.</li> <li>- Phân tích lỗi: Classification report.</li> </ul>	G2.1 G2.2	<b>Dạy:</b> GV hướng dẫn sinh viên tại lớp. <b>Học ở lớp:</b> Sinh viên thực hành theo hướng dẫn của giảng viên. <b>Học ở nhà:</b> Sinh viên hoàn thành bài thực hành theo yêu cầu.	A3
Buổi 6 (5 tiết)	<b>Bài thực hành 6:</b> Bài toán phân lớp ảnh. <ul style="list-style-type: none"> <li>- Bộ dữ liệu: Chest X-ray.</li> <li>- Các mô hình áp dụng: SVM, Logistic Regression.</li> <li>- Mạng neural.</li> <li>- Phân tích lỗi: Classification report.</li> </ul>	G2.1 G2.2	<b>Dạy:</b> GV hướng dẫn sinh viên tại lớp. <b>Học ở lớp:</b> Sinh viên thực hành theo hướng dẫn của giảng viên. <b>Học ở nhà:</b> Sinh viên hoàn thành bài thực hành theo yêu cầu.	A3

## 6. ĐÁNH GIÁ MÔN HỌC (Course assessment)

Thành phần đánh giá	CĐRMH	Tỷ lệ (%)
A1. Quá trình (Kiểm tra trên lớp, bài tập hằng tuần)	G1.1, G2.1	20%
A2. Giữa kỳ		0%
A3. Thực hành	G1.2, G2.1	30%
A4. Đồ án	G1.2, G2.1, G2.2	50%

**a. Rubric của thành phần đánh giá A1**

<b>Đánh giá quá trình học tập tại lớp</b>	<b>Giỏi (9-10đ)</b>	<b>Khá (7-8đ)</b>	<b>TB (6-7đ)</b>	<b>Yếu (4-5đ)</b>	<b>Kém (&lt;3đ)</b>
<i>Mức độ chuyên cần</i>	<i>Tích cực tham gia đóng góp và xây dựng các bài giảng. Chủ động trả lời câu hỏi và làm 100 %bài tập tại lớp.</i>	<i>Tham gia trả lời câu hỏi tại lớp và hoàn thành 100% bài tập tại lớp đầy đủ.</i>	<i>Hoàn thành 70% các bài tập tại lớp.</i>	<i>Hoàn thành 40% các bài tập tại lớp.</i>	<i>Không tham gia vào bất cứ hoạt động nào tại lớp.</i>

**b. Rubric của thành phần đánh giá A2**

Không có

**c. Rubric của thành phần đánh giá A3**

	<b>Giỏi (9-10đ)</b>	<b>Khá (7-8đ)</b>	<b>TB (6-7đ)</b>	<b>Yếu (4-5đ)</b>	<b>Kém (&lt;3đ)</b>
<i>Làm bài tập thực hành hằng tuần</i>	<i>Làm 100% bài tập</i>	<i>Làm 80% bài tập</i>	<i>Làm 70% bài tập</i>	<i>Làm dưới 50% bài tập</i>	<i>Không làm bài nào</i>
<i>Báo cáo thực hành</i>	<i>Nội dung đầy đủ, trình bày rõ ràng</i>	<i>Đáp ứng được 80% nội dung yêu cầu</i>	<i>Đáp ứng được 50% nội dung yêu cầu</i>	<i>Đáp ứng được 30% nội dung yêu cầu</i>	<i>Không đáp ứng hoặc sai nội dung yêu cầu</i>

**d. Rubric của thành phần đánh giá A4**

<b>Seminar/thuyết trình</b>	<b>Giỏi (9-10đ)</b>	<b>Khá (7-8đ)</b>	<b>TB (6-7đ)</b>	<b>Yếu (4-5đ)</b>	<b>Kém (&lt;3đ)</b>
<i>Nội dung</i>	<i>Nội dung đầy đủ, có lập luận tốt và minh họa rõ ràng</i>	<i>Nội dung đầy đủ, lập luận tương đối tốt và có minh họa rõ ràng</i>	<i>Nội dung tương đối đầy đủ, lập luận ổn.</i>	<i>Nội dung sơ sài</i>	<i>Sai nội dung hoặc không đáp ứng nội dung theo yêu cầu</i>
<i>Trình bày</i>	<i>Trình bày đẹp, bố cục hợp lý</i>	<i>Trình bày rõ ràng và bố cục hợp lý</i>	<i>Trình bày và bố cục đáp ứng theo yêu cầu.</i>	<i>Trình bày sơ sài.</i>	<i>Trình bày sai định dạng quy ước.</i>

## 7. QUY ĐỊNH CỦA MÔN HỌC (Course requirements and expectations)

- Sinh viên đến lớp học theo quy định chung của nhà trường.
- Sinh viên làm bài tập và thực hiện đồ án môn học đầy đủ.

## 8. TÀI LIỆU HỌC TẬP, THAM KHẢO

### Sách/giáo trình chính

1. Aurélien Géron (2017), *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O'Reilly Media Inc

### Tài liệu tham khảo

1. James Pustejovsky and Amber Stubbs (2012), *Natural Language Annotation for Machine Learning*, O'Reilly Media Inc.
2. Vũ Hữu Tiệp (2018), *Machine Learning Cơ bản*, NXB Khoa học và Kỹ thuật.

## 9. PHẦN MỀM HAY CÔNG CỤ HỖ TRỢ THỰC HÀNH

1. Python.
2. Google colab.

*Tp.HCM, ngày .... tháng .... năm .....*

**Trưởng khoa/bộ môn**

(Ký và ghi rõ họ tên)

**Giảng viên biên soạn**

(Ký và ghi rõ họ tên)