

**ĐỀ CƯƠNG MÔN HỌC**  
**IS252 – KHAI THÁC DỮ LIỆU**

**1. THÔNG TIN CHUNG**

Tên môn học (tiếng Việt):	<b>Khai thác dữ liệu</b>	
Tên môn học (tiếng Anh):	Data Mining	
Mã môn học:	IS252	
Thuộc khối kiến thức:	Chuyên ngành HTTT	
Khoa/Bộ môn phụ trách:	Hệ thống thông tin	
Website môn học		
Giảng viên phụ trách:	TS. Nguyễn Đình Thuân Email: thuannd@uit.edu.vn	
Giảng viên tham gia giảng dạy:	Ths. Mai Xuân Hùng Ths. Trịnh Minh Tuấn	
Số tín chỉ:	4	
	TC lý thuyết: 3	TC thực hành: 1
Lý thuyết: (tiết)	45	
Thực hành: (tiết)	30	
Tự học: (tiết)		
Tính chất của môn	Bắt buộc đối với sinh viên ngành HTTT	
Điều kiện đăng ký: (môn học tiên quyết, học trước, song hành...)	Xác suất thông kê, Cơ sở dữ liệu.	

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

Môn học nhằm cung cấp các kiến thức cơ bản về khai thác dữ liệu và quá trình khám phá tri thức, các giai đoạn chính của quá trình khai thác dữ liệu và khám phá tri thức.  
Sinh viên được trang bị các kỹ thuật chính trong khai thác dữ liệu như phân lớp, phân cụm, hồi quy, chuỗi thời gian, luật kết hợp...cũng như cách sử dụng các công cụ hỗ trợ xây dựng các ứng dụng khai thác dữ liệu.

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

Mục tiêu	Mô tả [1]	Mục tiêu (Theo CĐR cấp 3) [2]
G1	Kết hợp làm việc cá nhân và nhóm để thảo luận và thực	3.1.1, 3.1.2, 3.1.4,

	hiện đề tài theo nội dung môn học	3.2.1, 3.2.2
<b>G2</b>	+ Hiểu được khái niệm, sự cần thiết của việc khai thác dữ liệu. + Biết ứng dụng của việc khai thác dữ liệu trong các lĩnh vực của đời sống	1.3.6
<b>G3</b>	+ Hiểu được các bước trong quy trình khai thác dữ liệu + Khảo sát dữ liệu	1.3.6 2.2.2
<b>G4</b>	+ Xác định và phát biểu bài toán	2.1.1
<b>G5</b>	+ Xác định các chức năng của hệ thống	4.3.2
<b>G6</b>	+ Hiểu, cài đặt, thử nghiệm các thuật toán khai thác dữ liệu	1.2.7, 2.2.3
<b>G7</b>	Phân tích, đánh giá kết quả thuật toán so với kết quả dùng công cụ của SQL Server và Weka	2.1.4
<b>G8</b>	+ Thuyết trình + Đặt câu hỏi, lắng nghe và đối thoại	3.2.4 3.2.5

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

<b>CĐRMH [1]</b>	<b>CĐR cấp 4 của CTĐT [3]</b>	<b>Mô tả CĐRMH (mục tiêu cụ thể) [2]</b>	<b>Mức độ giảng dạy [4]</b>
G1.1	3.1.1.1	Xác định các thành phần nhóm, Xác định vai trò từng thành viên	U
G1.2	3.1.2.1	Xác định mục tiêu và những việc phải làm	U
G1.3	3.1.2.2	Lập lịch và địa điểm cho các cuộc họp nhóm	U
G1.4	3.1.2.3	Áp dụng các quy tắc làm việc nhóm	U
G1.5	3.1.2.4	Giao tiếp hiệu quả	U
G1.6	3.1.2.5	Hình thành các giải pháp cho các vấn đề được phân công	U
G1.7	3.1.2.6	Thể hiện tinh thần hợp tác, làm việc nhóm	U
G1.8	3.1.3.2	Duy trì và phát triển nhóm	U
G1.9	3.1.4.1	Giải thích mục tiêu, nhiệm vụ của nhóm và luôn định hướng trong quá trình hoạt động	U
G1.10	3.1.4.2	Lập kế hoạch thực hiện các công việc	U

G1.11	3.1.4.3	Xác định công việc cho từng thành viên phù hợp và thời hạn hoàn tất	U
G2.1	1.3.6.1	Hiểu được khái niệm khai thác dữ liệu là gì , sự cần thiết của việc khai thác dữ liệu.	T
G2.2	1.3.6.1	Sự cần thiết của việc khai thác dữ liệu trong đời sống thực tế.	T,U
G.2.3	1.3.6.1	Biết được việc ứng dụng các kỹ thuật khai thác dữ liệu trong các lĩnh vực khác nhau của đời sống.	T,U
G3.1	1.3.6.1	Hiểu được các bước trong quy trình khai thác dữ liệu.	T, U
G3.2	2.2.2.4	Hiểu và áp dụng bước tiền xử lý dữ liệu trong tiến trình khai phá dữ liệu	T,U
G4.1	2.1.1.1	Phân tích các dữ kiện cho từng kỹ thuật khai khác dữ liệu	T,U
G4.2	2.1.1.2	Lựa chọn bài toán giải quyết	T,U
G4.3	2.1.1.3	Mô tả bài toán	T,U
G5.1	4.3.2.1	Xác định các chức năng cần thiết của hệ thống	T,U
G5.2	4.3.2.2	Chọn các kỹ thuật công nghệ liên quan	U
G5.3	4.3.3.3	Phân rã hệ thống thành các thành phần chi tiết, chức năng của các, thành phần chi tiết và định nghĩa các giao diện	T, U
G6.1	1.2.7	Hiểu, vận dụng các thuật toán khai thác dữ liệu	T, U
G6.2	2.2.3.1	Cài đặt, chạy thử nghiệm các thuật toán khai thác dữ liệu trên các bộ dữ liệu chọn lựa	U
G7.1	2.1.4.1.	Phân tích kết quả các giải pháp khai thác dữ liệu với công cụ khai thác dữ liệu SQL Server và Weka	U, A
G7.2	2.1.4.2	Phân tích, đối chiếu đánh giá kết quả đạt được của các thuật toán khai thác dữ liệu với kết quả khi dùng công cụ khai thác dữ liệu của SQL Server và Weka	U, A
G8.1	3.2.4.2	Thuyết trình mật lạc, rõ ràng	U
G8.2	3.2.4.3	Sử dụng giao tiếp không lời	U

G8.3	3.2.5.1	Lắng nghe chủ động, tích cực trong các hoạt động tình huống giao tiếp	U
G8.4	3.2.5.2	Hình thành các câu hỏi và trả lời câu hỏi hiệu quả	U

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

### a. Lý thuyết:

Buổi (3 tiết) [1]	Nội dung	CĐRMH [3]	Hoạt động dạy và học [4]	Thành phần đánh giá
1	<b>Chương 1: Tổng quan về khai thác dữ liệu</b> 1.1. Quá trình khám phá tri thức 1.2. Các khái niệm 1.3. Ý nghĩa của việc khai thác dữ liệu 1.4. Ứng dụng của khai thác dữ liệu , trong từng lĩnh vực. 1.5. Chia nhóm thực hiện đề tài	G1.1, G1.3, G2.1, G2.2, G2.3	Dạy: Thuyết giảng, đặt câu hỏi cho sinh viên. Sinh viên thảo luận. Giải thích các câu hỏi, nêu ví dụ minh họa.	Các nhóm nêu ra một vài ứng dụng của việc khai thác dữ liệu trong thực tế
2	<b>Chương 2: Các vấn đề tiền xử lý dữ liệu</b> 2.1. Giới thiệu về tiền xử lý dữ liệu 2.2. Làm sạch dữ liệu 2.3. Tích hợp dữ liệu 2.4. Biến đổi dữ liệu 2.5. Thu giảm dữ liệu 2.6. Rời rạc hóa dữ liệu 2.7. Tạo cây phân cấp ý niệm 2.8. Biểu diễn dữ liệu	G4.1, G4.2	- Dạy : Thuyết giảng, hướng dẫn kỹ năng làm việc theo nhóm. Giảng viên tổng kết, đánh giá, kết luận - Học: Bài tập nhóm tại lớp.	Đánh giá dựa trên thảo luận và trình bày kết quả thực hiện bài tập nhóm trên lớp.
3,4	<b>Chương 3: Tập phổ biến và Luật kết hợp</b> 3.1. Tổng quan về luật kết hợp 3.2. Biểu diễn luật kết hợp 3.3. Khám phá các mẫu thường xuyên 3.4. Khám phá các kết hợp với giải thuật Apriori và các biến thể của giải thuật Apriori 3.5. Khám phá các kết hợp dựa trên tập phổ biến tối đại. 3.6. Cách tính độ tin cậy của luật	G4.3, G5.1, G5.2, G5.3, G6.1	- Dạy: Thuyết giảng, giảng viên đặt vấn đề. Giảng giải, giải thích-minh họa. Giảng viên tổng kết, đánh giá, kết luận. - Học: Sinh viên làm bài tập trên lớp (bài 1) và nộp	Đánh giá dựa trên thảo luận và trình bày kết quả thực hiện bài tập 1.
5	<b>Chương 4: Dãy phổ biến</b> 4.1. Tổng quan về dãy phổ biến 4.2. Cách tìm dãy phổ biến song song 4.3. Cách tìm dãy phổ biến tuần tự 4.4. Khám phá luật kết hợp từ dãy phổ biến, cách tính độ tin cậy của luật	G6.1, G6.2, G7.1, G7.2	- Dạy: Tổng kết bài tập trên lớp 1. Thuyết giảng, giảng viên đặt vấn đề. Giảng giải, giải thích-minh họa. Giảng viên tổng kết, đánh giá, kết luận. - Học: Sinh viên làm bài tập trên lớp	Đánh giá dựa trên thảo luận và trình bày kết quả thực hiện bài tập 2.

			(bài 2) và nộp	
6,7	<b>Chương 5: Tập thô</b> 5.1. Khái niệm tập thô 5.2. Quan hệ bất khả phân biệt 5.3. Xấp xỉ trên 5.4. Xấp xỉ dưới 5.5. Sự phụ thuộc thuộc tính 5.6. Thành lập ma trận phân biệt 5.7. Thiết lập hàm phân biệt từ ma trận phân biệt 5.8. Tìm các rút gọn từ hàm phân biệt	G6.1, G6.2, G7.1, G7.2	<ul style="list-style-type: none"> <li>- Dạy: Tổng kết bài tập trên lớp 2.</li> <li>- Thuyết giảng, giảng viên đặt vấn đề. Giảng giải, giải thích-minh họa. Giảng viên tổng kết, đánh giá, kết luận.</li> <li>- Học: Sinh viên làm bài tập trên lớp (bài 3) và nộp</li> </ul>	Đánh giá dựa trên thảo luận và trình bày kết quả thực hiện bài tập 3.
8,9	<b>Chương 6: Phân lớp dữ liệu</b> 6.1. Tổng quan về phân lớp dữ liệu 6.2. Phân lớp dữ liệu với cây quyết định 6.3. Phân lớp dữ liệu với mạng Bayesian 6.4. Phân lớp dữ liệu với mạng Neural	G6.1, G6.2, G7.1, G7.2	<ul style="list-style-type: none"> <li>- Dạy: Tổng kết bài tập trên lớp 3. Thuyết giảng, giảng viên đặt vấn đề. Giảng giải, giải thích-minh họa. Giảng viên tổng kết, đánh giá, kết luận.</li> <li>- Học: Sinh viên làm bài tập trên lớp 4 và nộp</li> </ul>	Đánh giá dựa trên thảo luận và trình bày kết quả thực hiện bài tập 4
10,11	<b>Chương 7: Góm cụm dữ liệu</b> 5.1. Tổng quan về gom cụm dữ liệu, các phương pháp gom cụm 5.2. Gom cụm dữ liệu bằng mạng Kmean 5.3. Gom cụm dữ liệu bằng mạng Kohonen 5.4. Các phương pháp gom cụm dữ liệu khác	G6.1, G6.2, G7.1, G7.2	<ul style="list-style-type: none"> <li>- Dạy: Tổng kết bài tập trên lớp 4. Thuyết giảng, giảng viên đặt vấn đề. Giảng giải, giải thích-minh họa. Giảng viên tổng kết, đánh giá, kết luận.</li> <li>- Học: Sinh viên làm bài tập trên lớp 5 và nộp</li> </ul>	Đánh giá dựa trên thảo luận và trình bày kết quả thực hiện bài tập 5
12,13	<b>Chương 8: Khai thác dữ liệu và công nghệ cơ sở dữ liệu</b> 8.1. Tổng quan về công nghệ cơ sở dữ liệu 8.2. Khả năng hỗ trợ khai thác dữ liệu của công nghệ cơ sở dữ liệu 8.3. Các ngôn ngữ truy vấn dành cho khai thác dữ liệu 8.4. Hỗ trợ của các DBMS ngày nay dành cho khai thác dữ liệu	G7.1, G7.2	<ul style="list-style-type: none"> <li>- Dạy: Tổng kết bài tập trên lớp 5. Thuyết giảng, giảng viên đặt vấn đề. Giảng giải, giải thích-minh họa. Giảng viên tổng kết, đánh giá, kết luận.</li> <li>- Học: Sinh viên làm bài tập trên lớp 6 và nộp</li> </ul>	Đánh giá dựa trên thảo luận và trình bày kết quả thực hiện bài tập 6
14	Báo cáo Project	G1.2, G1.3,	Sinh viên báo cáo,	Đánh giá dựa

		G1.4, G1.5, G1.6, G1.7, G1.8, G1.9, G1.10, G1.11, G8.1, G8.2, G8.3	các nhóm khác tham gia thảo luận, đánh giá Project của các nhóm khác Giáo viên: Đánh giá dựa vào chất lượng bài báo cáo, cách báo cáo, các công việc và mức độ hoàn thiện trong Project của từng thành viên trong nhóm.	trên việc trình bày báo cáo, trả lời câu hỏi, hình thức trình bày báo cáo. Giao diện và chức năng của chương trình.
15	<b>Ôn tập:</b> Tóm lược các nội dung chính cần phải nắm trong môn học	G8.4	<b>Dạy:</b> <i>giải đáp, nêu các hiểu nhầm thường gặp (common errors).</i> <b>Học ở lớp:</b> Sinh viên đặt câu hỏi	

### b. Thực hành

Buổi học (4 tiết) [1]	Nội dung [2]	CĐRMH [3]	Hoạt động dạy và học [4]	Hoạt động đánh giá
01	Nắm rõ các công cụ hỗ trợ bài toán khai phá dữ liệu	G4.2, G5.1, G5.3, G6.1	<p>❖ <u>Dạy:</u></p> <ul style="list-style-type: none"> <li>Hướng dẫn sử dụng các công cụ hỗ trợ bài toán khai phá dữ liệu: Analysis Services Multidimensional and Datamining, Weka Data Mining Software in Java.</li> </ul> <p>❖ <u>Học ở lớp:</u></p> <ul style="list-style-type: none"> <li>Sử dụng các chức năng của phần các phần mềm hỗ trợ</li> </ul> <p>❖ <u>Học ở nhà:</u></p> <ul style="list-style-type: none"> <li>Cài đặt và tìm hiểu các công cụ: Analysis Services Multidimensional and Datamining, Weka Data Mining Software in Java.</li> </ul>	Đánh giá dựa trên việc sử dụng các chức năng của phần mềm hỗ trợ
02,03	Thực hành được việc chuẩn bị tài nguyên cho bài toán khai phá dữ liệu. Bài toán luật kết hợp (Association rules)	G4.2, G5.1, G5.3, G6.1	<p>❖ <u>Dạy:</u></p> <ul style="list-style-type: none"> <li>Hướng dẫn cách tạo dựng data source, data source view, targeted mailing mining model structure.</li> <li>Xác định các data type và content type cho bài toán khai phá dữ liệu.</li> </ul>	Đánh giá dựa trên kết quả tìm tập phổ biến và luật kết hợp

			<ul style="list-style-type: none"> <li>– Khai phá luật kết hợp trên SQL Server.</li> <li>– Hướng dẫn cài đặt thuật toán tìm tập phổ biến và luật kết hợp</li> </ul> <p>❖ <u>Học ở lớp:</u></p> <ul style="list-style-type: none"> <li>– Sinh viên thực hiện theo hướng dẫn trong tài liệu tham khảo SQL Server 2012 Tutorials - Analysis Services Data Mining trên hệ thống (phần lesson 01)</li> </ul> <p>❖ <u>Học ở nhà:</u></p> <ul style="list-style-type: none"> <li>– Hoàn thành bài tập về nhà (BTVN01).</li> </ul>	
04,05	Thực hành được thao tác thêm và xử lý các mô hình trong khai phá dữ liệu.	G4.2, G5.1, G5.3, G6.1, G6.2	<p>❖ <u>Dạy:</u></p> <ul style="list-style-type: none"> <li>– Hướng dẫn các thao tác thêm mới clustering mining model, Naive Bayes mining model,..</li> </ul> <p>❖ <u>Học ở lớp:</u></p> <ul style="list-style-type: none"> <li>– Sinh viên thực hiện theo hướng dẫn trong tài liệu tham khảo SQL Server 2012 Tutorials - Analysis Services Data Mining trên hệ thống (phần lesson 03)</li> </ul> <p>❖ <u>Học ở nhà:</u></p> <ul style="list-style-type: none"> <li>– Hoàn thành bài tập về nhà (BTVN02).</li> </ul>	Đánh giá kết quả thực hiện bài tập về nhà 1
06,07	Thực hành được thao tác khai phá dữ liệu cho Exploring the Targeted Mailing Models	G4.2, G5.1, G5.3, G6.1, G6.2, G7.1, G7.2	<p>❖ <u>Dạy:</u></p> <ul style="list-style-type: none"> <li>– Thực hiện khai phá mô hình decision tree, clustering, naive bayes.</li> <li>– Thảo luận về bài tập về nhà 2, so sánh kết quả với các công cụ khai thác dữ liệu SQL Server và Weka</li> <li>– Hướng dẫn cài đặt bài tập về nhà 3</li> </ul> <p>❖ <u>Học ở lớp:</u></p> <ul style="list-style-type: none"> <li>– Sinh viên thực hiện theo hướng dẫn trong tài liệu tham khảo SQL Server 2012 Tutorials - Analysis Services Data Mining trên hệ thống (phần lesson 04)</li> </ul>	Đánh giá dựa trên kết quả thực hiện bài tập 2

			❖ Học ở nhà: – Hoàn thành bài tập về nhà (BTVN03).	
08	Thực hành được nội dung bài toán dự báo trong khai phá dữ liệu.	G4.2, G5.1, G5.3, G6.1, G6.2, G7.1, G7.2	❖ Dạy: – Thực hiện các thao tác cho bài toán dự báo (Predictions) – Thảo luận về bài tập về nhà 3 và so sánh kết quả đạt được với công cụ khai thác dữ liệu SQL Server và Weka ❖ Học ở lớp: – Sinh viên thực hiện theo hướng dẫn trong tài liệu tham khảo SQL Server 2012 Tutorials - Analysis Services Data Mining trên hệ thống (phần 05,06) ❖ Học ở nhà: – Hoàn chỉnh các báo cáo, hướng dẫn cài đặt và sử dụng cho các bài tập về nhà BTVN02, BTVN03.	Đánh giá kết quả thực hiện bài tập về nhà 3

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

Thành phần đánh giá [1]	CĐRMH (Gx) [2]	Tỷ lệ (%) [3]
Thảo luận nhóm, bài tập	G4, G5, G6, G7	10%
Thi lý thuyết giữa kỳ	G4, G6	20%
Đồ án môn học	G1, G2, G3, G4, G5, G6, G7, G8	30%
Thi lý thuyết cuối kỳ	G5, G6, G7, G8, G9, G10	40%

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

- Đăng ký và làm Project ở nhà mỗi nhóm gồm 2 đến 3 sinh viên
- Sinh viên lắng nghe khi thầy giảng và tham gia trao đổi, thảo luận, báo cáo trong quá trình làm việc nhóm trên lớp. Sinh viên phải đọc slide bài giảng của buổi học trước khi đến lớp, thực hiện nghiêm túc Project đã đăng ký cùng các bạn trong nhóm.
- Sinh viên phải hoàn thiện các bài thực hành ở phần thực hành
- Sinh viên phải tham gia từ 80% các buổi học trên lớp, phải tham gia báo cáo Project của nhóm.
- Hình thức thi giữa kỳ và cuối kỳ: Tự luận

## 8. TÀI LIỆU HỌC TẬP, THAM KHẢO

- [1]. **Đỗ Phúc**, Giáo trình + Slide Bài giảng Khai thác dữ liệu, ĐHQG TPHCM, 2005.  
[2]. **Hồ Tú Bảo**, Introduction to knowledge discovery and data mining, IOIT, 2001.  
[3]. **Jiawei Han and Micheline Kamber**, Data Mining Concepts and Techniques,



*University of Illinois, Morgan Kaufmann Publishers, 2006.*

[4]. **X. Wu, V. Kumar, J. Ross Quinlan, ...** *Top 10 Algorithms in Data Mining*, Chapman & Hall/CRC, Taylor & Francis Group, LLC, 2009.

[5]. **ZhaoHui Tang & Jamie MacLennan**, *Data Mining with SQL Server 2005*, Wiley Publishing, 2005.

[6]. Graham J. Williams, Simeon J. Simoff, “*Data Mining: Theory, Methodology, Techniques, and Applications*”, Springer-Verlag, 2006.

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

1. Trình quản trị CSDL SQL Server
2. Phần mềm WEKA

**Trưởng khoa/ bộ môn**

(Ký và ghi rõ họ tên)

**Giảng viên**

(Ký và ghi rõ họ tên)

## BÀI TẬP VỀ NHÀ 1

(BTVN1)

Môn học: Khai phá dữ liệu

### **Bài 1:**

Xét CSDL mẫu lưu trữ các hồ sơ tuyển dụng như sau:

**HỒ SƠ LÝ LỊCH TRÍCH NGANG**

Record_ID	Experienced_year	TOIEC	IELTS
ID001	0	450	5.5
ID002	6	400	4.5
ID003	7	500	8
ID004	0	650	6
ID005	0	700	5.5
ID006	2	550	4
ID007	6	500	4.5
ID008	1	350	4.5
ID009	5	420	5.5
ID010	2	480	6.5
ID011	6	470	7
ID012	2	380	8
ID013	8	500	9
ID014	7	550	5.5
ID015	3	600	6

**HỒ SƠ LÝ LỊCH TRÍCH NGANG**

Record_ID	Experienced_year	TOIEC	IELTS
ID016	1	700	4
ID017	2	650	4.5
ID018	3	640	5
ID019	8	630	5
ID020	8	550	5.5
ID021	5	400	7
ID022	7	450	6
ID023	0	490	6
ID024	7	480	6.5
ID025	2	650	4.5
ID026	7	700	6.5
ID027	8	600	7
ID028	3	450	7
ID029	1	390	5.5
ID030	2	470	8

### **Yêu cầu:**

- a. Xác định các hồ sơ được xét duyệt theo tiêu chí sau: số năm kinh nghiệm ít nhất là một năm

- và có chứng chỉ Anh văn TOEIC 500 điểm trở lên hoặc IELTS từ 4.5 trở lên.
- b.** Xác định tập luật kết hợp cho CSDL trên.

**Bài 2:**

Xét CSDL mẫu lưu trữ các hồ sơ tìm việc của sinh viên như trong file excel đính kèm

**Yêu cầu:**

- a.** Xác định các hồ sơ tìm việc theo các tiêu chí sau:
- Điểm rèn luyện  $\geq 90$  và điểm học tập trên 7.0 thì tìm việc dễ.
  - Điểm rèn luyện  $\geq 70$  và điểm học tập trên 6.0 thì tìm việc TB
  - Ngược lại thì tìm việc khó.
- b.** Xác định tập luật kết hợp cho CSDL trên

## BÀI TẬP VỀ NHÀ 2

(BTVN2)

Môn học: Khai phá dữ liệu

Viết chương trình khai phá luật kết hợp sử dụng thuật toán apriori

❖ Yêu cầu giao diện chương trình như sau:

The screenshot shows the Apriori software interface. It has a menu bar with 'File' and 'About'. The main area is divided into several sections:

- Items:** A text box for entering an item, with 'Add Item' and 'Delete Item' buttons.
- Item List:** A table with columns 'Item' and 'Support'. It contains items 'a', 'b', 'c', and 'd', all of which are checked.
- Transactions:** A table with columns 'Trans. Id' and 'Items'. It contains two transactions: '1' with items 'a,b' and '2' with items 'a,b,c'.
- min. Support:** A text box with the value '50' and a '%' sign.
- min. Confidence:** A text box with the value '50' and a '%' sign.
- Solve:** A button to execute the algorithm.
- Result:** A section containing two tables:
  - Frequent Items:** A table with columns 'Items' and 'Support'. It contains items 'a', 'b', 'c', and 'ab' with support values 4, 4, 2, and 4 respectively.
  - Strong Rules:** A table with columns 'Rules' and 'Confidence'. It contains rules 'a->b', 'b->a', 'a->c', 'c->a', and 'b->c' with confidence values 100.00%, 100.00%, 50.00%, 100.00%, and 50.00% respectively.

❖ Các chức năng của chương trình:

- Chức năng thêm thuộc tính (item): người dùng điền tên thuộc tính vào ô nhập *item*, sau đó chọn nút *Add Item*, thuộc tính mới tạo sẽ được hiển thị trong bảng *Item*.
- Chức năng xóa thuộc tính (item): người dùng chọn vào tên thuộc tính cần xóa trong bảng *Item*, sau đó chọn nút *Delete Item*, thuộc tính sẽ được xóa khỏi bảng *Item*.
- Chức năng thêm giao tác (transaction): người dùng muốn tạo giao tác gồm những thuộc tính nào thì chọn các thuộc tính trong bảng *Item*, sau đó chọn nút *Add Transaction*, giao tác mới tạo sẽ được thêm vào bảng *Transaction* bên dưới.
- Chức năng sửa giao tác (transaction): người dùng chọn vào tên giao tác cần sửa trong bảng *Transaction*, sau đó chọn nút *Edit Transaction* và điều chỉnh lại các thuộc tính cho giao tác đó, để hoàn tất việc sửa lại thuộc tính cho giao tác người dùng chọn nút *End Transaction*, và giao tác sẽ được cập nhật lại trong bảng *Transaction*.

- Chức năng xóa giao tác (transaction): người dùng chọn vào tên giao tác cần xóa trong bảng *Transaction*, sau đó chọn nút *Delete Transaction*, giao tác sẽ được xóa khỏi bảng *Transaction*.
- Sau khi thực hiện các chức năng trên, người dùng nhập *min support* và *min confidence* tương ứng vào các ô nhập.
- Người dùng chọn chức năng *Solve*, chương trình sẽ thực hiện ra kết quả tìm các tập phổ biến trong bảng *Frequent Items* và các luật kết hợp trong bảng *Strong Rules*.

### BÀI TẬP VỀ NHÀ 3

(BTVN3)

Môn học: Khai phá dữ liệu

Viết chương trình hiện thực cài đặt một thuật toán trong khai phá dữ liệu sau:

- Thuật toán cây quyết định
- Thuật toán gom cụm K-means
- Thuật toán Naïve Bayes
- ❖ Yêu cầu giao diện chương trình: thân thiện, dễ sử dụng, có hướng dẫn thao tác từng bước sử dụng, có các dữ liệu bài tập mẫu khi thực thi chương trình.
- ❖ Chức năng: thuật toán trong chương trình luôn cho kết quả đúng.

### ĐỒ ÁN MÔN HỌC

**ĐỒ ÁN 1:** Dùng thuật toán tìm tập phổ biến và luật kết hợp trong dữ liệu bán hàng điện tử qua mạng

**ĐỒ ÁN 2:** Dùng thuật toán chuỗi phổ biến để tìm tập phổ biến và luật kết hợp trong dữ liệu giá vàng

**ĐỒ ÁN 3:** Dùng kỹ thuật khai thác dữ liệu tập thô để tìm các rút gọn trong bài toán hỗ trợ ra quyết định trên dữ liệu tuyển sinh đại học 2015

**ĐỒ ÁN 4:** Gom cụm người tiêu dùng trong CSDL siêu thị bằng thuật toán gom cụm Kmean

**ĐỒ ÁN 5:** Dùng cây quyết định để phân lớp khách hàng vay vốn trong ngân hàng.

### THANG ĐIỂM BÁO CÁO ĐỒ ÁN

STT	Nội dung	Yêu cầu	Tỉ lệ điểm trong phần đánh giá Project
1	Chương trình demo	Đầy đủ chức năng	40%
2		Dữ liệu đầy đủ, nhiều	10%
		Giao diện thân thiện	5%
3	Cuốn báo cáo	Đầy đủ các phần	15%
4	Slide báo cáo	Đẹp, đơn giản, đầy đủ và lời cuốn	10%

5	Báo cáo	Đề hiểu, trả lời đúng những câu hỏi khi báo cáo	20%
---	---------	---	-----