

KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN KHOA HỌC MÁY TÍNH

Mã đề thi: **01-01**
Ngày thi: 31/12/2024

ĐỀ THI KẾT THÚC HỌC PHẦN

Tên học phần: **Khoa học dữ liệu**
Thời gian làm bài: 75 phút
Loại đề thi: **Tự luận**

Câu 1 (2.0 điểm): Hãy trình bày tổng quan về chu trình khai thác dữ liệu và các bước chính trong chu trình này.

Câu 2 (2.0 điểm): Nêu tổng quan về các thành phần trong hệ sinh thái Hadoop và vai trò của chúng.

Câu 3 (1.5 điểm): Thực hiện phép nhân chập ma trận (ảnh) đầu vào với ma trận hạt nhân sau với bước di chuyển ma trận hạt nhân trên ma trận đầu vào bằng 1, và không bổ sung thêm giá trị xung quang ma trận đầu vào.

Ma trận đầu vào:

3	1	3	0
5	4	6	1
9	8	7	2
3	1	2	1

Ma trận hạt nhân (3x3):

-1	0	1
0	-1	0
-1	0	1

Câu 4 (3.0 điểm): Cho bảng dữ liệu huấn luyện sau:

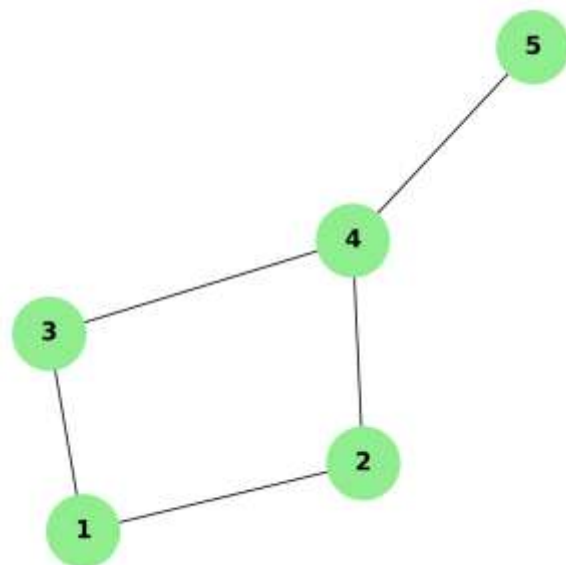
STT	Số giờ học	Số bài tập	Điểm kiểm tra
1	2	3	50
2	5	8	80
3	3	5	60
4	6	9	90
5	4	6	70

a. Xây dựng mô hình hồi quy tuyến tính dự đoán điểm kiểm tra dựa trên số giờ học và số bài tập mà học viên đã hoàn thành.

b. Tính trung bình lỗi tuyệt đối (MAE) của mô hình trên khi áp dụng vào dữ liệu kiểm tra sau:

STT	Số giờ học	Số bài tập	Điểm kiểm tra
1	5	7	85
2	4	6	75
3	3	4	60

Câu 5 (1.5 điểm): Hãy xác định đỉnh có độ đo trung tâm (betweenness centrality) cao nhất trong đồ thị dưới đây, và hãy giải thích lý do tại sao đỉnh đó có độ đo trung tâm cao nhất.



..... HẾT

Ghi chú: + Cán bộ coi thi không phải giải thích gì thêm
+ **Sinh viên không được sử dụng tài liệu**

Cán bộ ra đề
Nguyễn Hoàng Huy

Cán bộ duyệt đề
Trần Thị Thu Huyền

KHOA CÔNG NGHỆ THÔNG TIN
BỘ MÔN KHOA HỌC MÁY TÍNH
Mã đề thi: **01-02**
Ngày thi: 31/12/2024

ĐỀ THI KẾT THÚC HỌC PHẦN
Tên học phần: **Khoa học dữ liệu**
Thời gian làm bài: 75 phút
Loại đề thi: **Tự luận**

Câu 1 (2.0 điểm): Hãy nêu các lợi ích và thách thức chính của việc tích hợp dữ liệu từ nhiều nguồn khác nhau.

Câu 2 (2.0 điểm): Nêu tổng quan về Hadoop, mô tả kiến trúc của HDFS, và vai trò của NameNode và DataNode trong Hadoop.

Câu 3 (1.5 điểm): Thực hiện phép nhân chập ma trận (ảnh) đầu vào với ma trận hạt nhân sau với bước di chuyển ma trận hạt nhân trên ma trận đầu vào bằng 1, và không bổ sung thêm giá trị xung quang ma trận đầu vào.

Ma trận đầu vào:

1	2	3	0
4	5	6	1
7	8	9	2
0	1	2	3

Ma trận hạt nhân (3x3):

1	0	-1
0	1	0
1	0	-1

Câu 4 (3.0 điểm): Cho bảng dữ liệu huấn luyện sau:

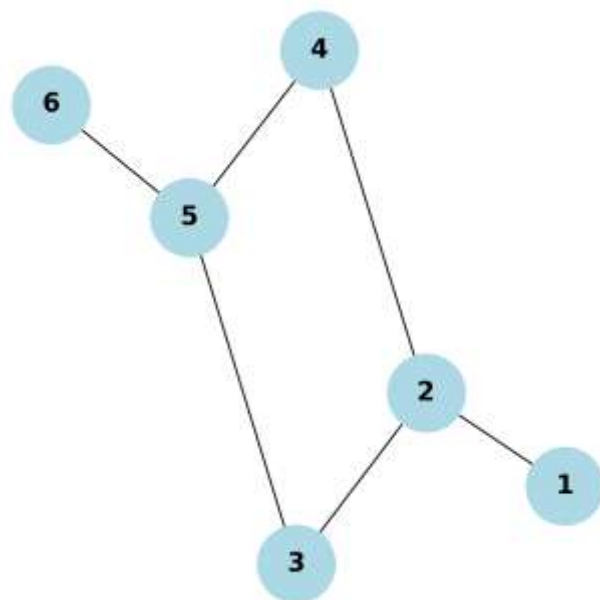
STT	Diện tích (m ²)	Số phòng	Giá nhà (triệu VND)
1	50	2	800
2	80	3	1200
3	60	2	950
4	100	4	1500
5	70	3	1100

a. Xây dựng mô hình hồi quy tuyến tính dự đoán giá nhà dựa trên diện tích và số phòng.

b. Tính trung bình lỗi tuyệt đối (MAE) của mô hình trên khi áp dụng vào dữ liệu kiểm tra sau:

STT	Diện tích (m ²)	Số phòng	Giá nhà (triệu VND)
1	90	3	1300
2	60	2	920
3	85	3	1250

Câu 5 (1.5 điểm): Hãy xác định đỉnh có độ đo trung tâm (betweenness centrality) cao nhất trong đồ thị dưới đây, và hãy giải thích lý do tại sao đỉnh đó có độ đo trung tâm cao nhất.



..... HẾT

Ghi chú: + Cán bộ coi thi không phải giải thích gì thêm

+ **Sinh viên không được sử dụng tài liệu**

Cán bộ ra đề
Nguyễn Hoàng Huy

Cán bộ duyệt đề
Trần Thị Thu Huyền