|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Danh sách lớp KHDV, học kỳ 1 năm học 2024-2025, học 13h55-17h40 Chiều thứ 6, Phòng 2-GĐ3. từ ngày 06/09/2024 – xx/xx/2024 (15 tuần)** | | | | | | | | | | | | | | | | | | | | | |
| *Stt* | *Mã SV* | *Họ và tên* | *Ngày sinh* | *Lớp* | *Buổi học tuần 1-15* | | | | | | | | | | | | | | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 0 | 1 | 2 | 3 | 4 | 5 | 6 |  |
|  | 22024545 | Hoàng Bảo An | 19/01/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  | Lê Quý Đôn, Đống Đa |
|  | 22024501 | Nguyễn Khắc An | 15/12/03 | K67T | × | × | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  | Hồng Thái ĐP HN |
|  | 22024530 | Đỗ Trần Vân Anh | 23/03/04 | K67T | × | × | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  | Chuyên Vĩnh Phúc |
|  | 22024564 | Lê Quốc Anh | 28/01/04 | K67T | × | × | M2 |  |  |  |  |  |  |  |  |  |  |  |  |  | Yên Hòa, |
|  | 22024542 | Nguyễn Duy Anh | 26/02/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  | Quang Trung HN |
|  | 22024512 | Nguyễn Duy Anh | 08/01/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  | Phạ Hồng Thái, BĐ HN |
|  | 22024508 | Trần Vỹ Anh | 03/02/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  | Chuyên KHTN |
|  | 22024524 | Ngô Ngọc Ánh | 28/04/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  | Chuyên Hòa Bình |
|  | 22024506 | Lê Xuân Bách | 10/01/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  | Ams HN |
|  | 22024525 | Phạm Thị Tùng Chi | 27/07/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024580 | Thái Thị Diệp | 29/01/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024554 | Nguyễn Tuấn Dũng | 25/08/04 | K67T | × | × | M |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024514 | Mạc Minh Duy | 30/10/04 | K67T | × | × | M |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024578 | Nguyễn Thế Duy | 25/12/04 | K67T | × | × | M |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024562 | Phạm Thế Duyệt | 09/12/04 | K67T | 1 | 1 | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024533 | Nguyễn Quý Dương | 18/01/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024518 | Nguyễn Tuấn Đạt | 25/05/04 | K67T | 1 | 1 | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024529 | Phan Tiến Đạt | 01/06/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024536 | Nguyễn Anh Đức | 03/01/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024561 | Phạm Văn Đức | 06/04/04 | K67T | × | × | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024572 | Phạm Hương Giang | 13/10/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024571 | Nguyễn Thị Thu Hà | 10/02/04 | K67T | 1 | 1 | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024576 | Đào Nguyên Hải | 13/11/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024532 | Nguyễn Đăng Hải | 12/07/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024534 | Nguyễn Tiến Việt Hải | 08/10/04 | K67T | × | × | O |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024535 | Đoàn Ngọc Hiếu | 07/11/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024517 | Hoàng Thu Hiếu | 09/03/04 | K67T | × | × | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024502 | Hồ Trung Hiếu | 14/01/04 | K67T | × | 1 | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024577 | Hoàng Đình Hoàn | 04/09/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024566 | Phan Đức Hùng | 10/03/04 | K67T | × | × | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024528 | Nguyễn Đức Huy | 21/07/04 | K67T | 1 | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024579 | Nguyễn Vũ Khánh Huy | 19/08/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024519 | Nguyễn Tuấn Hưng | 13/12/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024549 | Nguyễn Thị Hương | 25/08/04 | K67T | 2 | 1 | 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024521 | Lưu Quang Khải | 19/11/04 | K67T | × | 1 | 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024565 | Lê Văn Hoàng Khang | 02/12/04 | K67T | × | × | M |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024511 | Trần Minh Khanh | 02/01/04 | K67T | × | 1 | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024558 | Lương Gia Khánh | 25/04/04 | K67T | 1 | 1 | M |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024574 | Nguyễn Văn Kiên | 10/06/04 | K67T | 1 | × | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024546 | Lê Tuấn Kiệt | 27/07/04 | K67T | × | 1 | 2 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 20021012 | Nguyễn Ngọc Kỷ | 09/05/02 | K65T | × | × | O |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024516 | Nguyễn Thị Thanh Lam | 25/02/04 | K67T | 1 | × | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024513 | Lưu Quý Lân | 21/05/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024567 | Hoàng Linh | 09/08/04 | K67T | × | 1 | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024559 | Lê Hoàng Linh | 11/01/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024505 | Nguyễn Hà Linh | 05/09/04 | K67T | 1 | × | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024552 | Hà Đăng Long | 06/10/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024556 | Hoàng Bảo Long | 09/08/04 | K67T | × | × | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024539 | Vũ Hải Long | 10/09/00 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024551 | Hoàng Văn Lộc | 10/05/03 | K67T | × | × | O |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024522 | Trần Hoàng Lương | 07/07/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024515 | Hồ Nguyên Lượng | 13/08/04 | K67T | 1 | 1 | M1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024504 | Nguyễn Đức Mạnh | 21/11/04 | K67T | × | × | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024538 | Trần Hữu Mạnh | 13/04/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 20021392 | Lưu Đạt Tuấn Minh | 19/10/02 | K65T | × | × | m |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024540 | Nguyễn Đức Minh | 18/06/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024547 | Nguyễn Quang Minh | 22/04/04 | K67T | 1 | × | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024537 | Nguyễn Duy Nguyên | 17/05/04 | K67T | × | × | x |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024553 | Nguyễn Trung Nguyên | 11/07/03 | K67T | 1 | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024503 | Phùng Khôi Nguyên | 27/04/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 21021220 | Nguyễn Tuấn Nhật | 23/07/03 | K66T | × | × | O |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024573 | Nguyễn Yến Nhi | 20/08/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024569 | Nguyễn Thị Hồng Nhung | 08/04/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024568 | Nguyễn Đặng Nam Phong | 12/12/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024555 | Nguyễn Hoàng Phúc | 22/02/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024510 | Lê Ngọc Quang | 25/05/04 | K67T | 1 | 1 | p |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024531 | Phạm Tiến Sơn | 11/11/04 | K67T | 1 | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024500 | Lê Minh Tâm | 16/06/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024541 | Đỗ Tuấn Thành | 10/04/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024544 | Lê Đắc Thịnh | 06/05/04 | K67T | 1 | 1 | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024570 | Hồ Anh Thơ | 15/10/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024526 | Ngô Mạnh Tiến | 24/04/04 | K67T | × | × | O |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024575 | Đặng Sỹ Toàn | 20/09/04 | K67T | × | × | M |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024548 | Phạm Thu Trang | 06/06/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024520 | Lê Hồng Triệu | 04/10/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024560 | Đỗ Quang Trung | 22/05/04 | K67T | × | × | X |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024527 | Nguyễn Tiến Trung | 29/10/04 | K67T | × | × | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
|  | 22024523 | Nguyễn Thị Ánh Tuyết LT | 15/09/04 | K67T | 1 | × | 1 |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

Bạn Nguyến Thị Ánh Tuyết (MSSV 22024523) là lớp trưởng.

**Ghi chú:** “1”, “2”, “3”, …: số lần đóng góp bài trong buổi học, “x”: đi học đúng giờ, “×”: nhà trường chưa vào danh sách, (+): năm người đầu tiên; “M”: đi học muộn, “o”: bỏ học, “**Đ**”: sử dụng điện thoại trong giờ học, “**B**”: bỏ học giữa giờ.

Thời gian học kỳ 1 bắt đầu từ 08/02/2024.

Buổi 1. 06/09/2024 (Học trực tuyến): Chương 0. Giới thiệu môn học + Chương 1. Giới thiệu chung về HTTT

Buổi 2, 13/09/2024: Chương 2. HTTT trong Tổ chức

Buổi 3, 20/09/2024: Chương 3. Hiếu Thu Hiếu, Chi Phạm Tùng xin nghỉ tiết 4

Buổi 4, 27/09/2024:. Chương 4.

Buổi 5, 02/10/2024: Chương 5.

Buổi 6, 09/10/2024: Chương 6.

Buổi 7, 16/10/2024: Chương 6.

Buổi 8, 23/10/2024:. Chương 7.

Buổi 9, 30/10/2024: Chương 8.

Buổi 10, 05/11/2024: Chương 8.

Buổi 11-15, 17/11-06/05/2024: Các nhóm báo cáo tiểu luận

**Danh sách tiểu luận**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **MARKETING ANALYTICS  A Machine Learning Approach** | **1** | **Introduction to Marketing Analytics** | 22024527 | Nguyễn Tiến Trung | 18 |
| 22024544 | Lê Đắc Thịnh |
| 22024571 | Nguyễn Thị Thu Hà |
| 22024573 | Nguyễn Yến Nhi |
| **2** | **Statistics for Marketing** | 22024546 | Lê Tuấn Kiệt | 10 |
| 22024529 | Phan Tiến Đạt |
| 22024505 | Nguyễn Hà Linh |
| 22024545 | Hoàng Bảo An |
| **3** | **Evolution of Data Analytics** | 22024531 | Phạm Tiến Sơn | 4 |
| 22024552 | Hà Đăng Long |
| 22024502 | Hồ Trung Hiếu |
| 22024564 | Lê Quốc Anh |
| **4** | **Segmentation and Targeting Analysis** | 22024504 | Nguyễn Đức Mạnh | 9 |
| 22024503 | Phùng Khôi Nguyên |
| 22024528 | Nguyễn Đức Huy |
| 22024536 | Nguyễn Anh Đức |
| **5** | **Important Marketing Metrics:  A Snapshot** | 22024572 | Phạm Hương Giang | 2 |
| 22024530 | Đỗ Trần Vân Anh |
| 22024579 | Nguyễn Vũ Khánh Huy |
|  |  |
| **6** | **Consumer Buying Behavior** | 22024569 | Nguyễn Thị Hồng Nhung | 21 |
| 22024570 | Hồ Anh Thơ |
| 22024522 | Trần Hoàng Lương |
| 22024525 | Phạm Thị Tùng Chi |
| **7** | **Understanding Consumer Behavior  Using Market Basket Analysis** | 22024533 | Nguyễn Quý Dương | 19 |
| 22024559 | Lê Hoàng Linh |
| 22024535 | Đoàn Ngọc Hiếu |
| 22024574 | Nguyễn Văn Kiên |
| **8** | **Neuromarketing Techniques for  Consumer Analytics** | 22024500 | Lê Minh Tâm | 8 |
| 22024549 | Nguyễn Thị Hương |
| 22024517 | Hoàng Thu Hiếu |
| 22024548 | Phạm Thu Trang |
| **9** | **New Product Development** | 22024562 | Phạm Thế Duyệt | 11 |
| 22024515 | Hồ Nguyên Lượng |
| 22024520 | Lê Hồng Triệu |
| 22024521 | Lưu Quang Khải |
| **10** | **Natural Language Processing for  Branding** | 22024523 | Nguyễn Thị Ánh Tuyết | 22 |
| 22024537 | Nguyễn Duy Nguyên |
| 22024556 | Hoàng Bảo Long |
|  |  |
| **11** | **Forecasting Sales and Price** | 22024518 | Nguyễn Tuấn Đạt | 1 |
| 22024510 | Lê Ngọc Quang |
| 22024519 | Nguyễn Tuấn Hưng |
| 22024516 | Nguyễn Thị Thanh Lam |
| **12** | **Sales Prediction and Conversion** | 22024553 | Nguyễn Trung Nguyên | 23 |
| 22024547 | Nguyễn Quang Minh |
| 22024501 | Nguyễn Khắc An |
| 22024512 | Nguyễn Duy Anh |
| **13** | **Role of Supply Chain Analytics in  Marketing Analytics** | 22024580 | Thái Thị Diệp | 24 |
| 22024578 | Nguyễn Thế Duy |
| 22024524 | Ngô Ngọc Ánh |
| 22024577 | Hoàng Đình Hoàn |
| **14** | **Web and Social Media Analytics** | 22024511 | Trần Minh Khanh | 12 |
| 22024560 | Đỗ Quang Trung |
| 22024513 | Lưu Quý Lân |
| 22024514 | Mạc Minh Duy |
| **15** | **Marketing Analytics and  Its Applications** | 22024458 | Lương Gia Khánh | 15 |
| 22024540 | Nguyễn Đức Minh |
| 22024566 | Phan Đức Hùng |
| 22024567 | Hoàng Linh |
| **Information Systems Management: Governance, Urbanization and Alignment** | **1** | **IS Stakeholders** | 22024539 | Vũ Hải Long | 6 |
| 22024555 | Nguyễn Hoàng Phúc |
| 22024541 | Đỗ Tuấn Thành |
| 22024568 | Nguyễn Đặng Nam Phong |
| **2** | **From Global Governance  to IS Governance** | 22024508 | Trần Vỹ Anh | 7 |
| 22024506 | Lê Xuân Bách |
| 22024532 | Nguyễn Đăng Hải |
| 22024561 | Phạm Văn Đức |
| **3** | **IS Governance in Practice** | 22024538 | Trần Hữu Mạnh | 3 |
| 21021220 | Nguyễn Tuấn Nhật |
| 22024576 | Đào Nguyên Hải |
| 22024534 | Nguyễn Tiến Việt Hải |
| **4** | **The IS Territory** | 22024542 | Nguyễn Duy Anh | 5 |
| 20021012 | Nguyễn Ngọc Kỷ |
| 22024554 | Nguyễn Tuấn Dũng |
| 22024565 | Lê Văn Hoàng Khang |
| **5** | **Territorial Urbanization** |  |  |  |
|  |  |
|  |  |
|  |  |
| **6** | **Urbanizing the Inter-organizational IS** |  |  |  |
|  |  |
|  |  |
|  |  |
| **7** | **IS Project Management** |  |  |  |
|  |  |
|  |  |
|  |  |
| **8** | **Technology, Alignment and  Strategic Transformation** | 20021392 | Lưu Đạt Tuấn Minh | 13 |
| 22024551 | Hoàng Văn Lộc |
| 22024575 | Đặng Sỹ Toàn |
| 22024526 | Ngô Mạnh Tiến |
| **9** | **Auditing ISs** |  |  |  |
|  |  |
|  |  |
|  |  |

|  |  |
| --- | --- |
| **Sĩ số** | **78** |
| **Số nhóm** | **96** |
|  | |
| **Các bài báo đã được chọn** | |
| **1** | **TRUE** |
| **2** | **TRUE** |
| **3** | **TRUE** |
| **4** | **TRUE** |
| **5** | **TRUE** |
| **6** | **TRUE** |
| **7** | **TRUE** |
| **8** | **TRUE** |
| **9** | **TRUE** |
| **10** | **TRUE** |
| **11** | **TRUE** |
| **12** | **TRUE** |
| **13** | **TRUE** |
| **14** | **FALSE** |
| **15** | **TRUE** |
| **16** | **FALSE** |
| **17** | **FALSE** |
| **18** | **TRUE** |
| **19** | **TRUE** |
| **20** | **FALSE** |
| **21** | **TRUE** |
| **22** | **TRUE** |
| **23** | **TRUE** |
| **24** | **TRUE** |
| **25** | **FALSE** |
| **Chưa chọn** | **5** |

1. CSHTTT- Explainable. Arlind Kadra, Sebastian Pineda-Arango, Josif Grabocka. Breaking the Paradox of Explainable Deep Learning. CoRR abs/2305.13072 (2023). https://github.com/releaunifreiburg/inn; https://paperswithcode.com/paper/breaking-the-paradox-of-explainable-deep; https://www.catalyzex.com/paper/breaking-the-paradox-of-explainable-deep/code. Deep Learning has achieved tremendous results by pushing the frontier of automation in diverse domains. Unfortunately, current neural network architectures are not explainable by design. In this paper, we propose a novel method that trains deep hypernetworks to generate explainable linear models. Our models retain the accuracy of black-box deep networks while offering free lunch explainability by design. Specifically, our explainable approach requires the same runtime and memory resources as black-box deep models, ensuring practical feasibility. Through extensive experiments, we demonstrate that our explainable deep networks are as accurate as state-of-the-art classifiers on tabular data. On the other hand, we showcase the interpretability of our method on a recent benchmark by empirically comparing prediction explainers. The experimental results reveal that our models are not only as accurate as their black-box deep-learning counterparts but also as interpretable as state-of-the-art explanation techniques.
2. CSHTTT- Explainable. Chia-Yu Hsu, Wenwen Li. Explainable GeoAI: Can saliency maps help interpret artificial intelligence's learning process? An empirical study on natural feature detection. CoRR abs/2303.09660 (2023). https://github.com/asucicilab/explainable-geoai; https://paperswithcode.com/paper/explainable-geoai-can-saliency-maps-help; https://www.catalyzex.com/paper/explainable-geoai-can-saliency-maps-help/code. Improving the interpretability of geospatial artificial intelligence (GeoAI) models has become critically important to open the "black box" of complex AI models, such as deep learning. This paper compares popular saliency map generation techniques and their strengths and weaknesses in interpreting GeoAI and deep learning models' reasoning behaviors, particularly when applied to geospatial analysis and image processing tasks. We surveyed two broad classes of model explanation methods: perturbation-based and gradient-based methods. The former identifies important image areas, which help machines make predictions by modifying a localized area of the input image. The latter evaluates the contribution of every single pixel of the input image to the model's prediction results through gradient backpropagation. In this study, three algorithms-the occlusion method, the integrated gradients method, and the class activation map method-are examined for a natural feature detection task using deep learning. The algorithms' strengths and weaknesses are discussed, and the consistency between model-learned and human-understandable concepts for object recognition is also compared. The experiments used two GeoAI-ready datasets to demonstrate the generalizability of the research findings.
3. CSHTTT- Explainable. Felix Frohnert, Evert P. L. van Nieuwenburg. Explainable Representation Learning of Small Quantum States. CoRR abs/2306.05694 (2023). https://github.com/felixfrohnertdb/qcvae; https://paperswithcode.com/paper/explainable-representation-learning-of-small; https://www.catalyzex.com/paper/explainable-representation-learning-of-small/code. Unsupervised machine learning models build an internal representation of their training data without the need for explicit human guidance or feature engineering. This learned representation provides insights into which features of the data are relevant for the task at hand. In the context of quantum physics, training models to describe quantum states without human intervention offers a promising approach to gaining insight into how machines represent complex quantum states. The ability to interpret the learned representation may offer a new perspective on non-trivial features of quantum systems and their efficient representation. We train a generative model on two-qubit density matrices generated by a parameterized quantum circuit. In a series of computational experiments, we investigate the learned representation of the model and its internal understanding of the data. We observe that the model learns an interpretable representation which relates the quantum states to their underlying entanglement characteristics. In particular, our results demonstrate that the latent representation of the model is directly correlated with the entanglement measure concurrence. The insights from this study represent proof of concept towards interpretable machine learning of quantum states. Our approach offers insight into how machines learn to represent small-scale quantum systems autonomously.
4. CSHTTT- Explainable. Gabriele Dominici, Pietro Barbiero, Lucie Charlotte Magister, Pietro Liò, Nikola Simidjievski. SHARCS: Shared Concept Space for Explainable Multimodal Learning. CoRR abs/2307.00316 (2023). https://github.com/gabriele-dominici/SHARCS; https://paperswithcode.com/paper/sharcs-shared-concept-space-for-explainable; https://www.catalyzex.com/paper/sharcs-shared-concept-space-for-explainable/code. Multimodal learning is an essential paradigm for addressing complex real-world problems, where individual data modalities are typically insufficient to accurately solve a given modelling task. While various deep learning approaches have successfully addressed these challenges, their reasoning process is often opaque; limiting the capabilities for a principled explainable cross-modal analysis and any domain-expert intervention. In this paper, we introduce SHARCS (SHARed Concept Space) -- a novel concept-based approach for explainable multimodal learning. SHARCS learns and maps interpretable concepts from different heterogeneous modalities into a single unified concept-manifold, which leads to an intuitive projection of semantically similar cross-modal concepts. We demonstrate that such an approach can lead to inherently explainable task predictions while also improving downstream predictive performance. Moreover, we show that SHARCS can operate and significantly outperform other approaches in practically significant scenarios, such as retrieval of missing modalities and cross-modal explanations. Our approach is model-agnostic and easily applicable to different types (and number) of modalities, thus advancing the development of effective, interpretable, and trustworthy multimodal approaches.
5. CSHTTT- Explainable. Hanjing Wang, Dhiraj Joshi, Shiqiang Wang, Qiang Ji. Gradient-based Uncertainty Attribution for Explainable Bayesian Deep Learning. CoRR abs/2304.04824 (2023). https://www.catalyzex.com/paper/gradient-based-uncertainty-attribution-for/code. Predictions made by deep learning models are prone to data perturbations, adversarial attacks, and out-of-distribution inputs. To build a trusted AI system, it is therefore critical to accurately quantify the prediction uncertainties. While current efforts focus on improving uncertainty quantification accuracy and efficiency, there is a need to identify uncertainty sources and take actions to mitigate their effects on predictions. Therefore, we propose to develop explainable and actionable Bayesian deep learning methods to not only perform accurate uncertainty quantification but also explain the uncertainties, identify their sources, and propose strategies to mitigate the uncertainty impacts. Specifically, we introduce a gradient-based uncertainty attribution method to identify the most problematic regions of the input that contribute to the prediction uncertainty. Compared to existing methods, the proposed UA-Backprop has competitive accuracy, relaxed assumptions, and high efficiency. Moreover, we propose an uncertainty mitigation strategy that leverages the attribution results as attention to further improve the model performance. Both qualitative and quantitative evaluations are conducted to demonstrate the effectiveness of our proposed methods.
6. CSHTTT- Explainable. Mark Laurie, James Lu. Explainable Deep Learning for Tumor Dynamic Modeling and Overall Survival Prediction using Neural-ODE. CoRR abs/2308.01362 (2023). https://github.com/jameslu01/tdnode; https://paperswithcode.com/paper/explainable-deep-learning-for-tumor-dynamic; While tumor dynamic modeling has been widely applied to support the development of oncology drugs, there remains a need to increase predictivity, enable personalized therapy, and improve decision-making. We propose the use of Tumor Dynamic Neural-ODE (TDNODE) as a pharmacology-informed neural network to enable model discovery from longitudinal tumor size data. We show that TDNODE overcomes a key limitation of existing models in its ability to make unbiased predictions from truncated data. The encoder-decoder architecture is designed to express an underlying dynamical law which possesses the fundamental property of generalized homogeneity with respect to time. Thus, the modeling formalism enables the encoder output to be interpreted as kinetic rate metrics, with inverse time as the physical unit. We show that the generated metrics can be used to predict patients' overall survival (OS) with high accuracy. The proposed modeling formalism provides a principled way to integrate multimodal dynamical datasets in oncology disease modeling.
7. CSHTTT- Explainable. Pedro Sequeira, Melinda T. Gervasio. IxDRL: A Novel Explainable Deep Reinforcement Learning Toolkit based on Analyses of Interestingness. CoRR abs/2307.08933 (2023). https://github.com/sri-aic/23-xai-ixdrl-data; https://paperswithcode.com/paper/ixdrl-a-novel-explainable-deep-reinforcement; https://www.catalyzex.com/paper/ixdrl-a-novel-explainable-deep-reinforcement/code. In recent years, advances in deep learning have resulted in a plethora of successes in the use of reinforcement learning (RL) to solve complex sequential decision tasks with high-dimensional inputs. However, existing systems lack the necessary mechanisms to provide humans with a holistic view of their competence, presenting an impediment to their adoption, particularly in critical applications where the decisions an agent makes can have significant consequences. Yet, existing RL-based systems are essentially competency-unaware in that they lack the necessary interpretation mechanisms to allow human operators to have an insightful, holistic view of their competency. Towards more explainable Deep RL (xDRL), we propose a new framework based on analyses of interestingness. Our tool provides various measures of RL agent competence stemming from interestingness analysis and is applicable to a wide range of RL algorithms, natively supporting the popular RLLib toolkit. We showcase the use of our framework by applying the proposed pipeline in a set of scenarios of varying complexity. We empirically assess the capability of the approach in identifying agent behavior patterns and competency-controlling conditions, and the task elements mostly responsible for an agent's competence, based on global and local analyses of interestingness. Overall, we show that our framework can provide agent designers with insights about RL agent competence, both their capabilities and limitations, enabling more informed decisions about interventions, additional training, and other interactions in collaborative human-machine settings.
8. CSHTTT- Explainable. Yanci Zhang, Han Yu. LR-XFL: Logical Reasoning-based Explainable Federated Learning. CoRR abs/2308.12681 (2023). https://github.com/yanci87/lr-xfl; https://paperswithcode.com/paper/lr-xfl-logical-reasoning-based-explainable. Federated learning (FL) is an emerging approach for training machine learning models collaboratively while preserving data privacy. The need for privacy protection makes it difficult for FL models to achieve global transparency and explainability. To address this limitation, we incorporate logic-based explanations into FL by proposing the Logical Reasoning-based eXplainable Federated Learning (LR-XFL) approach. Under LR-XFL, FL clients create local logic rules based on their local data and send them, along with model updates, to the FL server. The FL server connects the local logic rules through a proper logical connector that is derived based on properties of client data, without requiring access to the raw data. In addition, the server also aggregates the local model updates with weight values determined by the quality of the clients' local data as reflected by their uploaded logic rules. The results show that LR-XFL outperforms the most relevant baseline by 1.19%, 5.81% and 5.41% in terms of classification accuracy, rule accuracy and rule fidelity, respectively. The explicit rule evaluation and expression under LR-XFL enable human experts to validate and correct the rules on the server side, hence improving the global FL model's robustness to errors. It has the potential to enhance the transparency of FL models for areas like healthcare and finance where both data privacy and explainability are important.
9. CSHTTT- Explainable. Zhongwei Yu, Jingqing Ruan, Dengpeng Xing. Explainable Reinforcement Learning via a Causal World Model. CoRR abs/2305.02749 (2023). https://github.com/easeonway/explainable-causal-reinforcement-learning; https://paperswithcode.com/paper/explainable-reinforcement-learning-via-a; https://www.catalyzex.com/paper/explainable-reinforcement-learning-via-a/code. Generating explanations for reinforcement learning (RL) is challenging as actions may produce long-term effects on the future. In this paper, we develop a novel framework for explainable RL by learning a causal world model without prior knowledge of the causal structure of the environment. The model captures the influence of actions, allowing us to interpret the long-term effects of actions through causal chains, which present how actions influence environmental variables and finally lead to rewards. Different from most explanatory models which suffer from low accuracy, our model remains accurate while improving explainability, making it applicable in model-based learning. As a result, we demonstrate that our causal model can serve as the bridge between explainability and learning.
10. CSHTTT- Explainable. Furkan Cantürk, Reyhan Aydogan. Explainable Active Learning for Preference Elicitation. CoRR abs/2309.00356 (2023). https://github.com/furkancanturk/explainable\_active\_learning; https://paperswithcode.com/paper/explainable-active-learning-for-preference; https://www.catalyzex.com/paper/explainable-active-learning-for-preference/code. Gaining insights into the preferences of new users and subsequently personalizing recommendations necessitate managing user interactions intelligently, namely, posing pertinent questions to elicit valuable information effectively. In this study, our focus is on a specific scenario of the cold-start problem, where the recommendation system lacks adequate user presence or access to other users' data is restricted, obstructing employing user profiling methods utilizing existing data in the system. We employ Active Learning (AL) to solve the addressed problem with the objective of maximizing information acquisition with minimal user effort. AL operates for selecting informative data from a large unlabeled set to inquire an oracle to label them and eventually updating a machine learning (ML) model. We operate AL in an integrated process of unsupervised, semi-supervised, and supervised ML within an explanatory preference elicitation process. It harvests user feedback (given for the system's explanations on the presented items) over informative samples to update an underlying ML model estimating user preferences. The designed user interaction facilitates personalizing the system by incorporating user feedback into the ML model and also enhances user trust by refining the system's explanations on recommendations. We implement the proposed preference elicitation methodology for food recommendation. We conducted human experiments to assess its efficacy in the short term and also experimented with several AL strategies over synthetic user profiles that we created for two food datasets, aiming for long-term performance analysis. The experimental results demonstrate the efficiency of the proposed preference elicitation with limited user-labeled data while also enhancing user trust through accurate explanations.
11. CSHTTT Task-Oriented Dialogue. Jinghong Chen, Weizhe Lin, Bill Byrne. Schema-Guided Semantic Accuracy: Faithfulness in Task-Oriented Dialogue Response Generation. CoRR abs/2301.12568 (2023). https://github.com/erichen0615/schemaguidedsemanticaccuracy; https://paperswithcode.com/paper/schema-guided-semantic-accuracy-faithfulness; https://www.catalyzex.com/paper/schema-guided-semantic-accuracy-faithfulness/code. Ensuring that generated utterances are faithful to dialogue actions is crucial for Task-Oriented Dialogue Response Generation. Slot Error Rate (SER) only partially measures generation quality in that it solely assesses utterances generated from non-categorical slots whose values are expected to be reproduced exactly. Utterances generated from categorical slots, which are more variable, are not assessed by SER. We propose Schema-Guided Semantic Accuracy (SGSAcc) to evaluate utterances generated from both categorical and non-categorical slots by recognizing textual entailment. We show that SGSAcc can be applied to evaluate utterances generated from a wide range of dialogue actions in the Schema Guided Dialogue (SGD) dataset with good agreement with human judgment. We also identify a previously overlooked weakness in generating faithful utterances from categorical slots in unseen domains. We show that prefix tuning applied to T5 generation can address this problem. We further build an ensemble of prefix-tuning and fine-tuning models that achieves the lowest SER reported and high SGSAcc on the SGD dataset.
12. CSHTTT- Task-Oriented Dialogue. Lang Cao. DiagGPT: An LLM-based Chatbot with Automatic Topic Management for Task-Oriented Dialogue. CoRR abs/2308.08043 (2023). https://www.catalyzex.com/paper/diaggpt-an-llm-based-chatbot-with-automatic/code. A significant application of Large Language Models (LLMs), like ChatGPT, is their deployment as chat agents, which respond to human inquiries across a variety of domains. While current LLMs proficiently answer general questions, they often fall short in complex diagnostic scenarios such as legal, medical, or other specialized consultations. These scenarios typically require Task-Oriented Dialogue (TOD), where an AI chat agent must proactively pose questions and guide users toward specific goals or task completion. Previous fine-tuning models have underperformed in TOD and the full potential of conversational capability in current LLMs has not yet been fully explored. In this paper, we introduce DiagGPT (Dialogue in Diagnosis GPT), an innovative approach that extends LLMs to more TOD scenarios. In addition to guiding users to complete tasks, DiagGPT can effectively manage the status of all topics throughout the dialogue development. This feature enhances user experience and offers a more flexible interaction in TOD. Our experiments demonstrate that DiagGPT exhibits outstanding performance in conducting TOD with users, showing its potential for practical applications in various fields.
13. CSHTTT- Task-Oriented Dialogue. Lingbo Mo, Shijie Chen, Ziru Chen, Xiang Deng, Ashley Lewis, Sunit Singh, Samuel Stevens, Chang-You Tai, Zhen Wang, Xiang Yue, Tianshu Zhang, Yu Su, Huan Sun. Roll Up Your Sleeves: Working with a Collaborative and Engaging Task-Oriented Dialogue System. CoRR abs/2307.16081 (2023). https://github.com/osu-nlp-group/tacobot; https://paperswithcode.com/paper/roll-up-your-sleeves-working-with-a; https://www.catalyzex.com/paper/roll-up-your-sleeves-working-with-a/code. We introduce TacoBot, a user-centered task-oriented digital assistant designed to guide users through complex real-world tasks with multiple steps. Covering a wide range of cooking and how-to tasks, we aim to deliver a collaborative and engaging dialogue experience. Equipped with language understanding, dialogue management, and response generation components supported by a robust search engine, TacoBot ensures efficient task assistance. To enhance the dialogue experience, we explore a series of data augmentation strategies using LLMs to train advanced neural models continuously. TacoBot builds upon our successful participation in the inaugural Alexa Prize TaskBot Challenge, where our team secured third place among ten competing teams. We offer TacoBot as an open-source framework that serves as a practical example for deploying task-oriented dialogue systems.
14. CSHTTT- Task-Oriented Dialogue. Namo Bang, Jeehyun Lee, Myoung-Wan Koo. Task-Optimized Adapters for an End-to-End Task-Oriented Dialogue System. CoRR abs/2305.02468 (2023). https://github.com/budzianowski/multiwoz; https://github.com/sogang-isds/TOATOD; https://www.catalyzex.com/paper/task-optimized-adapters-for-an-end-to-end/code. Task-Oriented Dialogue (TOD) systems are designed to carry out specific tasks by tracking dialogue states and generating appropriate responses to help users achieve defined goals. Recently, end-to-end dialogue models pre-trained based on large datasets have shown promising performance in the conversational system. However, they share the same parameters to train tasks of the dialogue system (NLU, DST, NLG), so debugging each task is challenging. Also, they require a lot of effort to fine-tune large parameters to create a task-oriented chatbot, making it difficult for non-experts to handle. Therefore, we intend to train relatively lightweight and fast models compared to PLM. In this paper, we propose an End-to-end TOD system with Task-Optimized Adapters which learn independently per task, adding only small number of parameters after fixed layers of pre-trained network. We also enhance the performance of the DST and NLG modules through reinforcement learning, overcoming the learning curve that has lacked at the adapter learning and enabling the natural and consistent response generation that is appropriate for the goal. Our method is a model-agnostic approach and does not require prompt-tuning as only input data without a prompt. As results of the experiment, our method shows competitive performance on the MultiWOZ benchmark compared to the existing end-to-end models. In particular, we attain state-of-the-art performance on the DST task of 2.2 dataset.
15. CSHTTT Task-Oriented Dialogue. Prajjwal Bhargava, Pooyan Amini, Shahin Shayandeh, Chinnadhurai Sankar. AUTODIAL: Efficient Asynchronous Task-Oriented Dialogue Model. CoRR abs/2303.06245 (2023). https://www.catalyzex.com/paper/autodial-efficient-asynchronous-task-oriented/code. As large dialogue models become commonplace in practice, the problems surrounding high compute requirements for training, inference and larger memory footprint still persists. In this work, we present AUTODIAL, a multi-task dialogue model that addresses the challenges of deploying dialogue model. AUTODIAL utilizes parallel decoders to perform tasks such as dialogue act prediction, domain prediction, intent prediction, and dialogue state tracking. Using classification decoders over generative decoders allows AUTODIAL to significantly reduce memory footprint and achieve faster inference times compared to existing generative approach namely SimpleTOD. We demonstrate that AUTODIAL provides 3-6x speedups during inference while having 11x fewer parameters on three dialogue tasks compared to SimpleTOD. Our results show that extending current dialogue models to have parallel decoders can be a viable alternative for deploying them in resource-constrained environments.
16. CSHTTT- Task-Oriented Dialogue. Qingyang Wu, James Gung, Raphael Shu, Yi Zhang. DiactTOD: Learning Generalizable Latent Dialogue Acts for Controllable Task-Oriented Dialogue Systems. CoRR abs/2308.00878 (2023). https://www.catalyzex.com/paper/diacttod-learning-generalizable-latent/code. Dialogue act annotations are important to improve response generation quality in task-oriented dialogue systems. However, it can be challenging to use dialogue acts to control response generation in a generalizable way because different datasets and tasks may have incompatible annotations. While alternative methods that utilize latent action spaces or reinforcement learning do not require explicit annotations, they may lack interpretability or face difficulties defining task-specific rewards. In this work, we present a novel end-to-end latent dialogue act model (DiactTOD) that represents dialogue acts in a latent space. DiactTOD, when pre-trained on a large corpus, is able to predict and control dialogue acts to generate controllable responses using these latent representations in a zero-shot fashion. Our approach demonstrates state-of-the-art performance across a wide range of experimental settings on the MultiWOZ dataset, including zero-shot, few-shot, and full data fine-tuning with both end-to-end and policy optimization configurations.
17. CSHTTT- Task-Oriented Dialogue. Stefania Raimondo, Christopher Pal, Xiaotian Liu, David Vázquez, Héctor Palacios. Improving Generalization in Task-oriented Dialogues with Workflows and Action Plans. CoRR abs/2306.01729 (2023). https://www.catalyzex.com/paper/improving-generalization-in-task-oriented/code. Task-oriented dialogue is difficult in part because it involves understanding user intent, collecting information from the user, executing API calls, and generating helpful and fluent responses. However, for complex tasks one must also correctly do all of these things over multiple steps, and in a specific order. While large pre-trained language models can be fine-tuned end-to-end to create multi-step task-oriented dialogue agents that generate fluent text, our experiments confirm that this approach alone cannot reliably perform new multi-step tasks that are unseen during training. To address these limitations, we augment the dialogue contexts given to \textmd{text2text} transformers with known \textit{valid workflow names} and \textit{action plans}. Action plans consist of sequences of actions required to accomplish a task, and are encoded as simple sequences of keywords (e.g. verify-identity, pull-up-account, reset-password, etc.). We perform extensive experiments on the Action-Based Conversations Dataset (ABCD) with T5-small, base and large models, and show that such models: a) are able to more readily generalize to unseen workflows by following the provided plan, and b) are able to generalize to executing unseen actions if they are provided in the plan. In contrast, models are unable to fully accomplish new multi-step tasks when they are not provided action plan information, even when given new valid workflow names.
18. CSHTTT- Task-Oriented Dialogue. Takyoung Kim, Jamin Shin, Young-Ho Kim, Sanghwan Bae, Sungdong Kim. Revealing User Familiarity Bias in Task-Oriented Dialogue via Interactive Evaluation. CoRR abs/2305.13857 (2023). https://www.catalyzex.com/paper/revealing-user-familiarity-bias-in-task/code. Most task-oriented dialogue (TOD) benchmarks assume users that know exactly how to use the system by constraining the user behaviors within the system's capabilities via strict user goals, namely "user familiarity" bias. This data bias deepens when it combines with data-driven TOD systems, as it is impossible to fathom the effect of it with existing static evaluations. Hence, we conduct an interactive user study to unveil how vulnerable TOD systems are against realistic scenarios. In particular, we compare users with 1) detailed goal instructions that conform to the system boundaries (closed-goal) and 2) vague goal instructions that are often unsupported but realistic (open-goal). Our study reveals that conversations in open-goal settings lead to catastrophic failures of the system, in which 92% of the dialogues had significant issues. Moreover, we conduct a thorough analysis to identify distinctive features between the two settings through error annotation. From this, we discover a novel "pretending" behavior, in which the system pretends to handle the user requests even though they are beyond the system's capabilities. We discuss its characteristics and toxicity while showing recent large language models can also suffer from this behavior.
19. CSHTTT- Task-Oriented Dialogue. Zhiyuan Hu, Yue Feng, Yang Deng, Zekun Li, See-Kiong Ng, Anh Tuan Luu, Bryan Hooi. Enhancing Large Language Model Induced Task-Oriented Dialogue Systems Through Look-Forward Motivated Goals. CoRR abs/2309.08949 (2023). https://www.catalyzex.com/paper/enhancing-large-language-model-induced-task/code. Recently, the development of large language models (LLMs) has been significantly enhanced the question answering and dialogue generation, and makes them become increasingly popular in current practical scenarios. While unlike the general dialogue system which emphasizes the semantic performance, the task-oriented dialogue (ToD) systems aim to achieve the dialogue goal efficiently and successfully in multiple turns. Unfortunately, existing LLM-induced ToD systems lack the direct reward toward the final goal and do not take account of the dialogue proactivity that can strengthen the dialogue efficiency. To fill these gaps, we introduce the ProToD (Proactively Goal-Driven LLM-Induced ToD) approach, which anticipates the future dialogue actions and incorporates the goal-oriented reward signal to enhance ToD systems. Additionally, we present a novel evaluation method that assesses ToD systems based on goal-driven dialogue simulations. This method allows us to gauge user satisfaction, system efficiency and successful rate while overcoming the limitations of current Information and Success metrics. Empirical experiments conducted on the MultiWoZ 2.1 dataset demonstrate that our model can achieve superior performance using only 10% of the data compared to previous end-to-end fully supervised models. This improvement is accompanied by enhanced user satisfaction and efficiency.
20. CSHTTT- Task-Oriented Dialogue. Weizhou Shen, Yingqi Gao, Canbin Huang, Fanqi Wan, Xiaojun Quan, Wei Bi. Retrieval-Generation Alignment for End-to-End Task-Oriented Dialogue System. CoRR abs/2310.08877 (2023). https://github.com/shenwzh3/mk-tod; https://paperswithcode.com/paper/retrieval-generation-alignment-for-end-to-end. Developing an efficient retriever to retrieve knowledge from a large-scale knowledge base (KB) is critical for task-oriented dialogue systems to effectively handle localized and specialized tasks. However, widely used generative models such as T5 and ChatGPT often struggle to differentiate subtle differences among the retrieved KB records when generating responses, resulting in suboptimal quality of generated responses. In this paper, we propose the application of maximal marginal likelihood to train a perceptive retriever by utilizing signals from response generation for supervision. In addition, our approach goes beyond considering solely retrieved entities and incorporates various meta knowledge to guide the generator, thus improving the utilization of knowledge. We evaluate our approach on three task-oriented dialogue datasets using T5 and ChatGPT as the backbone models. The results demonstrate that when combined with meta knowledge, the response generator can effectively leverage high-quality knowledge records from the retriever and enhance the quality of generated responses. The codes and models of this paper are available at this https URL: <https://github.com/shenwzh3/MK-TOD>.
21. CSHTTT\_ Fairness. Alexandra Kakadiaris. Evaluating the Fairness of the MIMIC-IV Dataset and a Baseline Algorithm: Application to the ICU Length of Stay Prediction. CoRR abs/2401.00902 (2024). https://github.com/akakadiaris/fairnessofmimiciv; https://paperswithcode.com/paper/evaluating-the-fairness-of-the-mimic-iv; https://www.catalyzex.com/paper/evaluating-the-fairness-of-the-mimic-iv/code. This paper uses the MIMIC-IV dataset to examine the fairness and bias in an XGBoost binary classification model predicting the Intensive Care Unit (ICU) length of stay (LOS). Highlighting the critical role of the ICU in managing critically ill patients, the study addresses the growing strain on ICU capacity. It emphasizes the significance of LOS prediction for resource allocation. The research reveals class imbalances in the dataset across demographic attributes and employs data preprocessing and feature extraction. While the XGBoost model performs well overall, disparities across race and insurance attributes reflect the need for tailored assessments and continuous monitoring. The paper concludes with recommendations for fairness-aware machine learning techniques for mitigating biases and the need for collaborative efforts among healthcare professionals and data scientists.
22. CSHTTT\_ Fairness. Ching-Hao Chiu, Yu-Jen Chen, Yawen Wu, Yiyu Shi, Tsung-Yi Ho. Achieve Fairness without Demographics for Dermatological Disease Diagnosis. CoRR abs/2401.08066 (2024). https://github.com/chiuhaohao/AttEN; https://paperswithcode.com/paper/achieve-fairness-without-demographics-for;. In medical image diagnosis, fairness has become increasingly crucial. Without bias mitigation, deploying unfair AI would harm the interests of the underprivileged population and potentially tear society apart. Recent research addresses prediction biases in deep learning models concerning demographic groups (e.g., gender, age, and race) by utilizing demographic (sensitive attribute) information during training. However, many sensitive attributes naturally exist in dermatological disease images. If the trained model only targets fairness for a specific attribute, it remains unfair for other attributes. Moreover, training a model that can accommodate multiple sensitive attributes is impractical due to privacy concerns. To overcome this, we propose a method enabling fair predictions for sensitive attributes during the testing phase without using such information during training. Inspired by prior work highlighting the impact of feature entanglement on fairness, we enhance the model features by capturing the features related to the sensitive and target attributes and regularizing the feature entanglement between corresponding classes. This ensures that the model can only classify based on the features related to the target attribute without relying on features associated with sensitive attributes, thereby improving fairness and accuracy. Additionally, we use disease masks from the Segment Anything Model (SAM) to enhance the quality of the learned feature. Experimental results demonstrate that the proposed method can improve fairness in classification compared to state-of-the-art methods in two dermatological disease datasets.
23. CSHTTT\_ Fairness. Junjie Yang, Jiajun Jiang, Zeyu Sun, Junjie Chen. A Large-scale Empirical Study on Improving the Fairness of Deep Learning Models. CoRR abs/2401.03695 (2024). https://github.com/junjie1003/DL-Fairness-Study; https://paperswithcode.com/paper/a-large-scale-empirical-study-on-improving. Fairness has been a critical issue that affects the adoption of deep learning models in real practice. To improve model fairness, many existing methods have been proposed and evaluated to be effective in their own contexts. However, there is still no systematic evaluation among them for a comprehensive comparison under the same context, which makes it hard to understand the performance distinction among them, hindering the research progress and practical adoption of them. To fill this gap, this paper endeavours to conduct the first large-scale empirical study to comprehensively compare the performance of existing state-of-the-art fairness improving techniques. Specifically, we target the widely-used application scenario of image classification, and utilized three different datasets and five commonly-used performance metrics to assess in total 13 methods from diverse categories. Our findings reveal substantial variations in the performance of each method across different datasets and sensitive attributes, indicating over-fitting on specific datasets by many existing methods. Furthermore, different fairness evaluation metrics, due to their distinct focuses, yield significantly different assessment results. Overall, we observe that pre-processing methods and in-processing methods outperform post-processing methods, with pre-processing methods exhibiting the best performance. Our empirical study offers comprehensive recommendations for enhancing fairness in deep learning models. We approach the problem from multiple dimensions, aiming to provide a uniform evaluation platform and inspire researchers to explore more effective fairness solutions via a set of implications.
24. CSHTTT\_ Fairness. Mike Laszkiewicz, Imant Daunhawer, Julia E. Vogt, Asja Fischer, Johannes Lederer. Benchmarking the Fairness of Image Upsampling Methods. CoRR abs/2401.13555 (2024). https://github.com/mikelasz/benchmarking-fairness-imageupsampling; https://paperswithcode.com/paper/benchmarking-the-fairness-of-image-upsampling; https://www.catalyzex.com/paper/benchmarking-the-fairness-of-image-upsampling/code. Recent years have witnessed a rapid development of deep generative models for creating synthetic media, such as images and videos. While the practical applications of these models in everyday tasks are enticing, it is crucial to assess the inherent risks regarding their fairness. In this work, we introduce a comprehensive framework for benchmarking the performance and fairness of conditional generative models. We develop a set of metrics–inspired by their supervised fairness counterparts–to evaluate the models on their fairness and diversity. Focusing on the specific application of image upsampling, we create a benchmark covering a wide variety of modern upsampling methods. As part of the benchmark, we introduce UnfairFace, a subset of FairFace that replicates the racial distribution of common large-scale face datasets. Our empirical study highlights the importance of using an unbiased training set and reveals variations in how the algorithms respond to dataset imbalances. Alarmingly, we find that none of the considered methods produces statistically fair and diverse results. All experiments can be reproduced using our provided repository.
25. CSHTTT\_ Fairness. Yibo Li, Xiao Wang, Yujie Xing, Shaohua Fan, Ruijia Wang, Yaoqi Liu, Chuan Shi. Graph Fairness Learning under Distribution Shifts. CoRR abs/2401.16784 (2024). https://www.catalyzex.com/paper/graph-fairness-learning-under-distribution/code. Graph neural networks (GNNs) have achieved remarkable performance on graph-structured data. However, GNNs may inherit prejudice from the training data and make discriminatory predictions based on sensitive attributes, such as gender and race. Recently, there has been an increasing interest in ensuring fairness on GNNs, but all of them are under the assumption that the training and testing data are under the same distribution, i.e., training data and testing data are from the same graph. Will graph fairness performance decrease under distribution shifts? How does distribution shifts affect graph fairness learning? All these open questions are largely unexplored from a theoretical perspective. To answer these questions, we first theoretically identify the factors that determine bias on a graph. Subsequently, we explore the factors influencing fairness on testing graphs, with a noteworthy factor being the representation distances of certain groups between the training and testing graph. Motivated by our theoretical analysis, we propose our framework FatraGNN. Specifically, to guarantee fairness performance on unknown testing graphs, we propose a graph generator to produce numerous graphs with significant bias and under different distributions. Then we minimize the representation distances for each certain group between the training graph and generated graphs. This empowers our model to achieve high classification and fairness performance even on generated graphs with significant bias, thereby effectively handling unknown testing graphs. Experiments on real-world and semi-synthetic datasets demonstrate the effectiveness of our model in terms of both accuracy and fairness.