PHÂN TÍCH CÁC ĐỀ TÀI

Đề tài 1: Tree RNN: Tree Structural Recurrent Neural Network for Network Traffic Classification

* Tập trung về mảng: An ninh mạng, phân loại lưu lượng mạng.
* Độ khó khi tìm hiểu: Cao (Hiểu rõ về cấu trúc cây và cách áp dụng RNN vào dữ liệu không tuần tự.
* Tài liệu tìm hiểu: Khá hạn chế.

Đề tài 2: Nghiên cứu các phương pháp phát hiện tin giả

* Tập trung về mảng: Kiểm duyệt nội dung, truyền thông.
* Độ khó khi tìm hiểu: Vừa vừa. Có nhiều phương pháp tìm hiểu đơn giản như SVM hay NB đến phức tạp như RNN, Transformer.
* Khả năng tìm hiểu đề tài: Nguồn tài liệu dồi dào (chẳng hạn như tin tức, các bài viết giả mạo). Nhiều thư viện hỗ trợ như scilet-learn, TensorFlow.

Đề tài 3: Application of Data Mining Algorithms for Dementia in People with HIV/AIDS

* Ứng dụng thực tiễn: Y học, chuẩn đoán.
* Độ khó tìm hiểu: Vừa vừa, chủ yếu tập trung vào các thuật toán khai thác dữ liệu cơ bản (PCA, Logistic Regression).
* Khả năng tìm hiểu đề tài: Cần dữ liệu y tế chuyên biệt, có thể khó tiếp cận đối với CNTT.

Đề tài 4: DEGAN: Decentralized Generative Adversarial Networks

* Ứng dụng thực tiễn: Blockchain, IoT, hệ thống phi tập trung (Hệ thống phi tập trung là một loại hệ thống trong đó các thành phần không phụ thuộc vào một điểm kiểm soát hoặc quản lý duy nhất.)
* Độ khó khi tìm hiểu: Cao, cần tìm hiểu sâu về GAN (Mạng đối kháng sinh) và hệ thống phi tập trung
* Khả năng tìm hiểu đề tài: Phù hợp với nghiên cứu chuyên sâu nhưng không lý tưởng cho sinh viên ở cấp độ “gà”.

Đề tài 5. Two-step Classification with SVD Pre-Processing of Distributed Massive Datasets in Apache Spark

* Ứng dụng thực tiễn: Big Data, phân loại dữ liệu lớn.
* Độ phức tạp: Cao, yêu cầu hiểu sâu về Spark và tiền xử lý với SVD.
* Khả năng tìm hiểu: Đòi hỏi cơ sở hạ tầng lớn là gì

Đề tài 6: Các phương pháp trích chọn đặc trưng trong hệ thống nhận dạng vân tay

* Ứng dụng thực tiễn: Sinh trắc học.
* Độ khó khi tìm hiểu: Vừa vừa, sương sương. Thường liên quan đến hình ảnh.
* Khả năng khai thác: Tài liệu phổ biến, ứng dụng thực tế rõ ràng.

Đề tài 7. Nghiên cứu hệ thống chia sẻ xe phi tập trung an toàn bằng Blockchain

* Ứng dụng thực tiễn: Thường thì liên quan đến chia sẻ phương tiện, bảo mật thông tin.
* Độ khó khi tìm hiểu đề tài: Cao, yêu cầu hiểu về blockchain và hợp đồng thông minh.
* Khả năng khai thác: Kiến thức nâng cao về blockchain.

Đề tài 8. Contents-based Recommendation Systems

* Ứng dụng thực tiễn: Liên quan đến gợi ý sản phẩm, dịch vụ (chẳng hạn như Netflix, Spotify).
* Độ phức tạp: Easy (các phương pháp như cosine similarity hoặc matrix factorization không quá phức tạp)
* Khả năng khai thác đề tài: Dữ liệu và thư viện dễ tiếp cận, phù hợp cho sinh viên trình độ “phường/khóm”.

Đề tài 9. Các kỹ thuật xác thực dựa trên mống mắt

* Ứng dụng thực tiễn: Xác thực sinh trắc học.
* Độ khó: Trung bình, sương sương. yêu cầu xử lý ảnh và trích xuất đặc trưng.
* Khả năng khai thác: Phổ biến nhưng cần dữ liệu ảnh mống mắt.

Đề tài 10: EpiGraphDB: a database and data mining platform for health data science

* Ứng dụng thực tiễn: Lĩnh vực y tế (dữ liệu về sức khỏe này nọ).
* Độ phức tạp: Cao ngút trời luôn, cần phải tìm hiểu hệ thống và nền tảng kiến thức EpiGraphDB.
* Khả năng khai thác đề tài: Phù hợp với nghiên cứu chuyên sâu.

Đề tài 11: Large Graph Sampling Algorithm For Frequent Subgraph Mining

* Ứng dụng thực tiễn: Liên quan đến vấn đề đồ thị lớn trong mạng xã hội hoặc mạng dữ liệu.
* Độ phức tạp: Cao luôn, Một số thuật toán phức tạp như RASI.
* Khả năng khai thác: Đề tài hẹp, khó thực hiện (Ý kiến riêng: Next).

Đề tài 12: Artificial Neural Network-based Techniques for Anomaly Detection on Apache Spark

* Ứng dụng thực tiễn: Phát hiện bất thường trong hệ thống lớn.
* Độ phức tạp khi tìm hiểu: Trung bình, có thể triển khai với Spark MLlib và TensorFlow.
* Khả năng khai thác đề tài: Yêu cầu cơ sở hạ tầng phân tán, nhưng phù hợp cho ứng dụng thực tế.

Đề tài 13: A Data Mining Framework for Financial Prediction

* Ứng dụng thực tiễn: Phân tích tài chính, dự đoán giá cổ phiếu.
* Độ khó khi tìm hiểu: Vừa vừa, tập trung vào khai thác dữ liệu và mô hình dự đoán.
* Khả năng khai thác đề tài: Tài liệu phong phú, nhiều vô số kể.

Đề tài 14: Distributed Frequent Subgraph Mining on Evolving Graph Using Spark

* Ứng dụng thực tiễn: Phân tích dữ liệu đồ thị lớn, cập nhật động.
* Độ khó khi tìm hiểu: Cao, yêu cầu hiểu Spark và cơ chế xử lý đồ thị động.
* Khả năng khai thác: Khó cho sinh viên nếu không có nền tảng về đồ thị. (Next)

Đề tài 15: A Parallel Algorithm for Graph Transaction Based Frequent Subgraph Mining

* Ứng dụng thực tiễn: Phân tích đồ thị giao dịch (transactional graphs).
* Độ phức tạp: Cao, yêu cầu nền tảng song song hóa và khai thác đồ thị con.
* Khả năng khai thác: Thích hợp hơn cho nghiên cứu sau đại học.

Tất cả các chủ đề: Code đều có (Python)

Đối với sinh viên đại học (dễ triển khai, tài liệu sẵn có):

* Nghiên cứu các phương pháp phát hiện tin giả.
* Contents-based Recommendation Systems.
* Các phương pháp trích chọn đặc trưng trong hệ thống nhận dạng vân tay.

Đối với sinh viên muốn thử thách và có nền tảng tốt: (Sinh viên trình độ “VIP”)

* Artificial Neural Network-based Techniques for Anomaly Detection on Apache Spark.
* A Data Mining Framework for Financial Prediction.
* Tree RNN for Network Traffic Classification.