TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**Ngô Chí Cường - 52100778**

**Lê Trần Phú – 52100572**

**USING GRAPH PATTERN MINING**

**FOR ANALYZING OF**

**MAKING MAXIMUM PROFIT**

**IN SELL CULTURE**

**DỰ ÁN CÔNG NGHỆ THÔNG TIN**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

TỔNG LIÊN ĐOÀN LAO ĐỘNG VIỆT NAM

**TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

**KHOA CÔNG NGHỆ THÔNG TIN**



**Ngô Chí Cường - 52100778**

**Lê Trần Phú – 52100572**

**USING GRAPH PATTERN MINING**

**FOR ANALYZING OF**

**MAKING MAXIMUM PROFIT**

**IN SELL CULTURE**

**DỰ ÁN CÔNG NGHỆ THÔNG TIN**

Người hướng dẫn

**ThS Dzoãn Xuân Thanh**

**THÀNH PHỐ HỒ CHÍ MINH, NĂM 2024**

**LỜI CẢM ƠN**

Trước tiên, chúng em xin gửi tới thầy Dzoãn Xuân Thanh lời cảm ơn chân thành nhất. Thầy là người đã mang đến cho chúng em nhiều kiến thức, kinh nghiệm và trong lúc đồng hành để hoàn thành Dự Án Công Nghệ Thông Tin. Suốt quá trình đồng hành cùng chúng em, Thầy luôn đưa ra những kiến thức chuyên môn sâu rộng, đầy sự tận tâm và nhiệt huyết.

Chúng em cũng xin bày tỏ lòng biết ơn sâu sắc đến tất cả các giáo viên trong Khoa Công nghệ thông tin của Trường Đại học Tôn Đức Thắng. Nhờ những kiến thức và kinh nghiệm mà các thầy cô đã chia sẻ, chúng em đã có thể hoàn thiện bài báo cáo này với chất lượng tốt nhất.

Cuối cùng, chúng em xin gửi lời cảm ơn đến gia đình và các bạn trong nhóm thực hành của chúng em. Những người bạn này đã luôn ở bên chúng em trong suốt quá trình học tập và giúp đỡ chúng em, là nguồn động lực rất lớn giúp chúng em tiến bộ hơn trong môn học của mình.

Chúng em rất cảm kích và tự hào khi được học tập và trải nghiệm những điều tuyệt vời như vậy. Chúng em hy vọng rằng sẽ có thể tiếp tục nhận được sự giúp đỡ và hướng dẫn từ các thầy cô để phát triển bản thân và đóng góp vào xã hội. Một lần nữa, chúng em xin chân thành cảm ơn!

*TP. Hồ Chí Minh, ngày 26 tháng 5 năm 2024*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

*Ngô Chí Cường*

*Lê Trần Phú*

**CÔNG TRÌNH ĐƯỢC HOÀN THÀNH**

**TẠI TRƯỜNG ĐẠI HỌC TÔN ĐỨC THẮNG**

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi và được sự hướng dẫn khoa học của Ths Dzoãn Xuân Thanh. Các nội dung nghiên cứu, kết quả trong đề tài này là trung thực và chưa công bố dưới bất kỳ hình thức nào trước đây. Những số liệu trong các bảng biểu phục vụ cho việc phân tích, nhận xét, đánh giá được chính tác giả thu thập từ các nguồn khác nhau có ghi rõ trong phần tài liệu tham khảo.

Ngoài ra, trong Dự án còn sử dụng một số nhận xét, đánh giá cũng như số liệu của các tác giả khác, cơ quan tổ chức khác đều có trích dẫn và chú thích nguồn gốc.

**Nếu phát hiện có bất kỳ sự gian lận nào tôi xin hoàn toàn chịu trách nhiệm về nội dung Dự án của mình**. Trường Đại học Tôn Đức Thắng không liên quan đến những vi phạm tác quyền, bản quyền do tôi gây ra trong quá trình thực hiện (nếu có).

*TP. Hồ Chí Minh, ngày 26 tháng 5 năm 2024*

*Tác giả*

*(Ký tên và ghi rõ họ tên)*

*Ngô Chí Cường*

*Lê Trần Phú*

**CLASSTIFICATION & REGESSION AND OVERFITTING**

**TÓM TẮT**

Đề tài của chúng em sẽ trình bày về lý thuyết của việc Graph Mining dựa trên việc phân tích dữ liệu tĩnh không thay đổi theo thời gian và dữ liệu động được thay đổi theo thời gian và áp dụng các thuật toán xử lý dữ liệu thành các đồ thị con từ đó phân tích như TKG,AER… Vừa củng cố các cơ sở lý thuyết đã nêu và đồng thời thể hiện tính ứng dụng của Graph Mining vào cuộc sống

**MỤC LỤC**

[DANH MỤC HÌNH VẼ vii](#_Toc167656630)

[DANH MỤC BẢNG BIỂU x](#_Toc167656631)

[DANH MỤC CÁC CHỮ VIẾT TẮT xi](#_Toc167656632)

[CHƯƠNG 1:NEURAL NETWORK 1](#_Toc167656633)

[1.1 Các optimizer khác nhau trong huấn luyện Neural Network 1](#_Toc167656634)

[1.1.1 Stochastic Gradient Descent (SGD): 1](#_Toc167656635)

[1.1.2 Momentum: 1](#_Toc167656636)

[1.1.3 Adagrad: 2](#_Toc167656637)

[1.1.4 RMSprop 3](#_Toc167656638)

[1.1.5 Adam: 4](#_Toc167656639)

[1.2 Ưu và nhược điểm các optimizer 5](#_Toc167656640)

[1.3 Áp dụng vào bài toán và so sánh kết quả 7](#_Toc167656641)

[1.3.1 Chuẩn bị thư viện. 7](#_Toc167656642)

[1.3.2 Chuẩn bị dữ liệu 7](#_Toc167656643)

[1.3.3 Xây dựng mô hình 7](#_Toc167656644)

[1.3.4 Khởi tạo Optimizer 8](#_Toc167656645)

[1.3.5 Huấn luyện mô hình với từng Optimizer 8](#_Toc167656646)

[1.3.6 Kết quả huấn luyện với các optimizer. 9](#_Toc167656647)

[1.3.7 So sánh và kết luận 11](#_Toc167656648)

[CHƯƠNG 2. LONG SHORT TERM MEMORY 14](#_Toc167656649)

[2.1 Nguyên nhân phát triển 14](#_Toc167656650)

[2.2 Khái niệm cơ bản 14](#_Toc167656651)

[2.3 Cấu trúc và cách hoạt động 14](#_Toc167656652)

[2.4 Các biến thể khác của LSTM 17](#_Toc167656653)

[2.4.1 GRU - Gated Recurrent Unit 17](#_Toc167656654)

[2.4.2 LSTM 2 chiều- Bidirectional LSTM 17](#_Toc167656655)

[2.5 Ứng dụng LSTM 18](#_Toc167656656)

[2.6 Demo mô hình LSTM cho bài toán dự đoán dữ liệu có đặc điểm time-series 19](#_Toc167656657)

[2.6.1 Import thư viện 19](#_Toc167656658)

[2.6.2 Chuẩn bị dữ liệu 19](#_Toc167656659)

[2.6.3 Thao tác tạo và huấn luyện mô hình 21](#_Toc167656660)

[2.6.4 Đánh giá trung bình bình phương của sai số của tập dữ liệu huấn luyện và tập dữ liệu xác thực 22](#_Toc167656661)

[2.6.5 Tiến hành dự đoán bằng mô hình và kết quả dự đoán 22](#_Toc167656662)

[2.6.6 Đánh giá giá trị MSE của tập dự đoán so với kết quả thực tế của mô hình LSTM 23](#_Toc167656663)

[2.7 So sánh với các mô hình học máy khác 23](#_Toc167656664)

[2.7.1 LinearRegresion 23](#_Toc167656665)

[2.7.2 RidgeRegression 25](#_Toc167656666)

[2.8 Áp dụng các biện pháp tránh OverFitting cho mô hình 27](#_Toc167656667)

[2.8.1 Thao tác tạo và huấn luyện mô hình và áp dụng các biện pháp tránh Overfitting 27](#_Toc167656668)

[2.8.2 Đánh giá trung bình bình phương của sai số của tập dữ liệu huấn luyện và tập dữ liệu xác thực 27](#_Toc167656669)

[2.8.3 Tiến hành so sánh dự đoán 28](#_Toc167656670)

[2.8.4 Điểm sai số trung bình giữa dữ liệu thử nghiệm và dữ liệu mô hình đoán được 28](#_Toc167656671)

[2.9 Các phương pháp cải tiến mô hình 29](#_Toc167656672)

[2.9.1 Sử dụng tối ưu siêu tham số (Hyperparameter Optimization) 29](#_Toc167656673)

[2.9.2 Sử dụng Feature Selection 30](#_Toc167656674)

[2.10 Ứng dụng các phương pháp trên để cải tiến mô hình 30](#_Toc167656675)

[2.10.1 Sử dụng tối ưu siêu tham số (Hyperparameter Optimization) 30](#_Toc167656676)

[2.10.2 Sử dụng Feature Selection 33](#_Toc167656677)

[CHƯƠNG 3. PHÂN CÔNG CÔNG VIỆC 38](#_Toc167656678)

[CHƯƠNG 4. TÀI LIỆU THAM KHẢO 39](#_Toc167656679)

[4.1 Tiếng Việt 39](#_Toc167656680)

[4.2 Tiếng Anh 39](#_Toc167656681)

# DANH MỤC HÌNH VẼ

[Hình 1.3.1.1:Hình ảnh đoạn code bước thêm thư viện cho Neural Network 7](#_Toc167656682)

[Hình 1.3.2.1.2: Hình ảnh đoạn code tải bộ dữ liệu và chia thành tập huấn luyện và tập kiểm tra 7](#_Toc167656683)

[Hình 1.3.2.2.3: Hình ảnh đoạn code tiền xử lý dữ liệu 7](#_Toc167656684)

[Hình 1.3.3.4: Hình ảnh đoạn code xây dựng mô hình 8](#_Toc167656685)

[Hình 1.3.4.5: Hình ảnh đoạn code khởi tạo Optimizer 8](#_Toc167656686)

[Hình 1.3.5.6: Đoạn code huấn luyện từng mô hình với Optimizer 9](#_Toc167656687)

[Hình 1.3.6.1.7: Kết quả huấn luyện mô hình với SGD Optimizer 9](#_Toc167656688)

[Hình 1.3.6.2.8: Kết quả huấn luyện mô hình với Momentum Optimizer 10](#_Toc167656689)

[Hình 1.3.6.3.9: Kết quả huấn luyện mô hình với Adagrad Optimizer 10](#_Toc167656690)

[Hình 1.3.6.4.10: Kết quả huấn luyện mô hình với RMSprop Optimizer 11](#_Toc167656691)

[Hình 1.3.6.5.11: Kết quả huấn luyện mô hình với Adam optimizer 11](#_Toc167656692)

[Hình 1.3.7.1.12: Hình ảnh đồ thị biểu hiện độ chính xác của các Optimizer khi sử dụng vào 1 mô hình và 1 tập dữ liệu 12](#_Toc167656693)

[Hình 2.3.13: Hình ảnh cấu trúc của một mô hình LSTM 15](#_Toc167656694)

[Hình 2.3.14: Hình ảnh cấu trúc của một Forget Gate 15](#_Toc167656695)

[Hình 2.3.15: Hình ảnh cấu trúc của một Input Gate 16](#_Toc167656696)

[Hình 2.3.16: Hình ảnh việc mô hình kết hợp những kết quả đầu ra 16](#_Toc167656697)

[Hình 2.3.17: Hình ảnh cấu trúc của một Output Gate 17](#_Toc167656698)

[Hình 2.4.1.18: Hình ảnh mô hình GRU 17](#_Toc167656699)

[Hình 2.4.2.19: Hình ảnh của mô hình LSTM 2 chiều 18](#_Toc167656700)

[Hình 2.6.1.20: Hình ảnh đoạn code thêm thư viện cho LSTM 19](#_Toc167656701)

[Hình 2.6.2.1.21: Hình ảnh đoạn code đọc dữ liệu cổ phiếu 20](#_Toc167656702)

[Hình 2.6.2.2.22: Hình ảnh sử dụng thư viện MinMaxScaler để giảm độ lớn dữ liệu 20](#_Toc167656703)

[Hình 2.6.2.2.23: Hình ảnh biến dữ liệu thành dạng time-series 21](#_Toc167656704)

[Hình 2.6.2.4.24: Hình ảnh chia dữ liệu thành tập huấn luyện và tập dự đoán 21](#_Toc167656705)

[Hình 2.6.2.4.25: Hình ảnh tạo mô hình LSTM và thực hiện huấn luyện trên ma trận giá trị Closed 21](#_Toc167656706)

[Hình 2.6.4.26: Hình ảnh đồ thị biểu hiện giá trị MSE qua từng tập huấn luyện 22](#_Toc167656707)

[Hình 2.6.5.27: Hình ảnh kết quả dự đoán của mô hình LSTM 23](#_Toc167656708)

[Hình 2.6.6.28: Hình ảnh giá trị MSE của tập dự đoán so với kết quả thực tế của mô hình LSTM 23](#_Toc167656709)

[Hình 2.7.1.1.29: Hình ảnh tạo mô hình LinearRegression 24](#_Toc167656710)

[Hình 2.7.1.2.30: Hình ảnh kết quả dự đoán của mô hình LinearRegresion 24](#_Toc167656711)

[Hình 2.7.1.3.31: Hình ảnh giá trị MSE của tập dự đoán so với kết quả thực tế của mô hình LSTM 25](#_Toc167656712)

[Hình 2.7.2.1.32: Hình ảnh tạo mô hình RidgeRegression 25](#_Toc167656713)

[Hình 2.7.2.2.33: Hình ảnh kết quả dự đoán của mô hình LinearRegresion 26](#_Toc167656714)

[Hình 2.7.2.3.34: Giá trị MSE của tập dự đoán so với kết quả thực tế của mô hình RidgeRegression 26](#_Toc167656715)

[Hình 2.8.1.35: Tạo mô hình LSTM với 2 biện pháp tránh OverFitting là Dropout Dense và EarlyStopping 27](#_Toc167656716)

[Hình 2.8.2.36: Hình ảnh so sánh đồ thị giá trị MSE qua từng tập huấn luyện 27](#_Toc167656717)

[Hình 2.8.3.37: Hình ảnh so sánh kết quả dự đoán của 2 mô hình so với giá trị thực tế 28](#_Toc167656718)

[Hình 2.8.3.38: Hình ảnh giá trị MSE của 2 mô hình có method chống Overfitting và không có method chống Overfitting 29](#_Toc167656719)

[Hình 2.10.1.1.39: Hình ảnh thực hiện huấn luyện mô hình có tối ưu siêu tham số 31](#_Toc167656720)

[Hình 2.10.1.2.40: Hình ảnh đoạn code so sánh mô hình có tối ưu siêu tham số và không có tối ưu siêu tham số thông qua đồ thị 32](#_Toc167656721)

[Hình 2.10.1.2.41: Hình ảnh đồ thị so sánh giá trị MSE mô hình có tối ưu siêu tham số và không có tối ưu siêu tham số thông qua đồ thị 32](#_Toc167656722)

[Hình 2.10.1.3.42: Hình ảnh đoạn code và so sánh kết quả dự đoán của 2 mô hình 33](#_Toc167656723)

[Hình 2.10.2.1.43: Hình ảnh thực hiện chuẩn bị Feature Selection 34](#_Toc167656724)

[Hình 2.10.2.2.44: Thực hiện huấn luyện mô hình với các Feature Selection đã chọn 35](#_Toc167656725)

[Hình 2.10.2.3.45: Hình ảnh đồ thị so sánh giá trị MSE mô hình có tối ưu siêu tham số và không có tối ưu siêu tham số thông qua đồ thị 36](#_Toc167656726)

[Hình 2.10.2.4.46: Hình ảnh kết quả dự đoán của 2 mô hình 36](#_Toc167656727)

# DANH MỤC BẢNG BIỂU

[Bảng 1.2.1: Bảng so sánh ưu điểm và nhược điểm các optimzer của mô hình Neural Network 7](#_Toc167656728)

[Bảng 3.2: Bảng phân công công việc 38](#_Toc167656729)

# DANH MỤC CÁC CHỮ VIẾT TẮT

|  |  |
| --- | --- |
| MSE | Mean Squared Error |
| LSTM | Long Short Term Memory |
|  |  |
|  |  |
|  |  |
|  |  |

# **MỞ ĐẦU**

# **CƠ SỞ LÝ THUYẾT**

# **PHÂN TÍCH THIẾT KẾ**

# TÀI LIỆU THAM KHẢO

## Tiếng Việt

< Đại Học Tôn Đức Thắng, (2024), Slide bài giảng nhập môn học máy>

**<** Nttuan8, (2019), Long short term memory (LSTM), <https://nttuan8.com/bai-14-long-short-term-memory-lstm/> >

<phamdinhkhanh, (2019), Lý thuyết về mạng LSTM part 2, <https://phamdinhkhanh.github.io/2019/04/22/Ly_thuyet_ve_mang_LSTM.html> />

<HCMUS, (2022), HCMUS - Summer - AI - Machine Learning part 2 - RNN, LSTM, <https://www.youtube.com/watch?v=PJKtY4WWwbQ> />

## Tiếng Anh

< StatsQuest, (2023), Long Short-Term Memory (LSTM), Clearly Explained, <https://www.youtube.com/watch?v=YCzL96nL7j0&t=485s> / >

< Grey Hogg, (2022) , LSTM Time Series Forecasting Tutorial in Python, https://www.youtube.com/watch?v=c0k-YLQGKjY&t=259s/ >

<Shipra Saxena, (2024), What is LSTM? Introduction to Long Short-Term Memory, <https://www.analyticsvidhya.com/blog/2021/03/introduction-to-long-short-term-memory-lstm/> >

< Colah, (2015), Understanding LSTM Networks, <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> >

< Kaggle, (2021), Microsoft Stock, <https://www.kaggle.com/datasets/vijayvvenkitesh/microsoft-stock-time-series-analysis> [/](https://scikit-learn.org/stable/supervised_learning.html#supervised-learning/)>