|  |  |
| --- | --- |
| HỌC VIỆN NÔNG NGHIỆP VIỆT NAM  **KHOA CÔNG NGHỆ THÔNG TIN** | CỘNG HOÀ XÃ HỘI CHỦ NGHĨA VIỆT NAM  **Độc lập - Tự do - Hạnh phúc** |

****

**ĐỀ CƯƠNG KHÓA LUẬN TỐT NGHIỆP**

**Đề tài: Phân lớp dữ liệu dự đoán lỗi phần mềm**

**Dựa trên thuật toán SVM**

**Họ và Tên sinh viên: Nguyễn Thị Duyên**

**Mã sinh viên: 637614**

**Lớp: Lớp K63HTTT**

**Giảng viên hướng dẫn: Th.S Nguyễn Xuân Thảo**

**HÀ NỘI - 2024**

**MỤC LỤC**

[PHẦN 1: MỞ ĐẦU 3](#_Toc159524556)

[1.1. Đặt vấn đề 3](#_Toc159524557)

[1.2. Mục đích và yêu cầu 3](#_Toc159524558)

[Phần 2: Tổng quan tình hình nghiên cứu trog và ngoài nước 4](#_Toc159524559)

[2.1. Tình hình nghiên cứu trong nước 4](#_Toc159524560)

[2.2. Tình hình nghiên cứu ngoài nước 4](#_Toc159524561)

[2.3. Tên đề tài, tính thời sự và tầm quan trọng của đề tài 4](#_Toc159524562)

[Phần 3: Nội dung và phương pháp nghiên cứu 5](#_Toc159524563)

[3.1. Địa điểm và thời gian thực tập 5](#_Toc159524564)

[3.2. Nội dung nghiên cứu 6](#_Toc159524565)

[3.3. Phương pháp nghiên cứu 6](#_Toc159524566)

[Phần 4: Kế hoạch và dự kiến kết quả đạt được 6](#_Toc159524567)

[Phần 5: Tài liệu tham khảo 7](#_Toc159524568)

# **PHẦN 1: MỞ ĐẦU**

* 1. **Đặt vấn đề**

Trong kỷ nguyên công nghệ số hiện nay, việc đảm bảo chất lượng phần mềm là một yếu tố quan trọng không chỉ đối với sự thành công của các doanh nghiệp công nghệ mà còn đối với an toàn và hiệu quả của hệ thống thông tin toàn cầu. Khi mà phần mềm ngày càng trở nên phức tạp và rộng lớn, việc phát hiện và sửa chữa lỗi phần mềm sớm trở thành một thách thức lớn, đặc biệt là trong các giai đoạn đầu của quá trình phát triển phần mềm. Điều này đòi hỏi phải có một phương pháp hiệu quả và chính xác để phân tích và dự đoán khả năng xuất hiện lỗi trong các dự án phần mềm.

Thuật toán Máy Vector Hỗ Trợ (SVM) là một trong những phương pháp học máy mạnh mẽ, đã được chứng minh là có hiệu quả trong nhiều lĩnh vực khác nhau, từ phân loại ảnh đến nhận dạng giọng nói và dự đoán thị trường tài chính. Trong lĩnh vực phát triển phần mềm, SVM hứa hẹn sẽ mở ra một hướng tiếp cận mới cho việc phân lớp và dự đoán lỗi phần mềm, giúp các nhà phát triển có thể chủ động ngăn chặn và giảm thiểu rủi ro gặp phải trong quá trình phát triển.

Vì vậy, đề tài "Phân lớp dữ liệu dự đoán lỗi phần mềm dựa trên thuật toán SVM" được chọn làm đề tài thực tập tốt nghiệp với mục tiêu nghiên cứu và ứng dụng thuật toán SVM trong việc phân loại và dự đoán lỗi phần mềm. Đề tài này không chỉ có ý nghĩa trong việc nâng cao chất lượng và độ tin cậy của phần mềm mà còn góp phần vào việc tối ưu hóa quy trình phát triển phần mềm, giảm thiểu chi phí và thời gian cần thiết cho việc kiểm thử và sửa chữa lỗi.

* 1. **Mục đích và yêu cầu**

**1.2.1 Mục đích**

* Nắm vững cơ sở lý thuyết của thuật toán SVM (Máy Vector Hỗ Trợ), bao gồm nguyên lý hoạt động, cách thức phân lớp dữ liệu và ứng dụng của nó trong phân loại và dự đoán.
* Phát triển một mô hình phân lớp dữ liệu dự đoán lỗi phần mềm dựa trên thuật toán SVM, cho phép phân biệt giữa các mẫu dữ liệu có khả năng cao và thấp xảy ra lỗi.
* Tìm hiểu và phân tích các ứng dụng thực tế của SVM trong lĩnh vực phát triển phần mềm, như dự đoán lỗi, phân loại yêu cầu người dùng, và tối ưu hóa quá trình kiểm thử phần mềm.

**1.2.2 Yêu cầu**

**Bài toán:** Cho văn bản tiếng Anh hoặc tiếng Việt. Thực hiện mã hóa và giải mã văn bản trên bằng thuật toán mã hóa RSA.

* Lý thuyết:
  + Tìm hiểu sâu rộng về lý thuyết phân lớp và dự đoán trong học máy, đặc biệt là thuật toán SVM và cách thức ứng dụng nó để giải quyết các vấn đề thực tế.
  + Tìm hiểu về cách thức SVM phân biệt và phân lớp dữ liệu, cũng như các kỹ thuật tối ưu hóa và lựa chọn mô hình.
* Chương trình: Phát triển một chương trình sử dụng ngôn ngữ lập trình Python để cài đặt mô hình SVM cho việc phân lớp dữ liệu và dự đoán lỗi phần mềm.
  + Tiền xử lý dữ liệu: Chuẩn bị và làm sạch dữ liệu đầu vào để phù hợp với mô hình SVM.
  + Xây dựng mô hình SVM: Sử dụng thư viện học máy như scikit-learn để xây dựng và huấn luyện mô hình.
  + Đánh giá mô hình: Sử dụng các phương pháp đánh giá như cross-validation và các chỉ số đánh giá (precision, recall, F1-score) để đánh giá hiệu suất mô hình.
  + Ứng dụng mô hình: Áp dụng mô hình đã được huấn luyện để phân lớp và dự đoán lỗi trong các dự án phần mềm.

# **Phần 2: Tổng quan tình hình nghiên cứu trong và ngoài nước**

## **2.1. Tình hình nghiên cứu trong nước**

* Ở Việt Nam, việc ứng dụng và nghiên cứu về thuật toán SVM trong lĩnh vực công nghệ thông tin và phần mềm đang dần được mở rộng. Các trường đại học và tổ chức nghiên cứu đã bắt đầu chú trọng vào việc phát triển các mô hình dự đoán và phân lớp dữ liệu, nhất là trong lĩnh vực dự đoán lỗi phần mềm.
* Các nghiên cứu chủ yếu tập trung vào việc tối ưu hóa các thuật toán SVM để nâng cao độ chính xác và hiệu quả trong việc phân loại và dự đoán. Điều này bao gồm việc điều chỉnh các tham số, lựa chọn tính năng, và sử dụng các kỹ thuật tiền xử lý dữ liệu hiệu quả.
* Một số công trình nghiên cứu cũng đề cập đến việc áp dụng SVM trong các ứng dụng cụ thể như phân tích chất lượng mã nguồn, dự đoán khả năng bảo trì phần mềm, và nhận dạng các mẫu thiết kế phần mềm dễ gặp lỗi.

## **2.2. Tình hình nghiên cứu ngoài nước**

* Trên phạm vi quốc tế, nghiên cứu về thuật toán SVM trong dự đoán lỗi phần mềm là một lĩnh vực rộng lớn với nhiều công trình đa dạng. Các nhà nghiên cứu đã khám phá nhiều phương pháp mới để cải thiện khả năng dự đoán của SVM, bao gồm việc phát triển các phiên bản SVM nâng cao và kết hợp với các thuật toán học máy khác.
* Một số nghiên cứu quốc tế tập trung vào việc áp dụng SVM trong các môi trường phát triển phần mềm cụ thể, như phát triển Agile hoặc DevOps, và trong các dự án sử dụng các công nghệ mới như microservices và containers..
* Nghiên cứu cũng mở rộng vào việc tích hợp SVM với các công nghệ mới nổi như trí tuệ nhân tạo (AI), machine learning deep learning, blockchain, và Internet of Things (IoT) để tạo ra các giải pháp toàn diện hơn trong việc dự đoán và quản lý lỗi phần mềm.

**Tổng quan này nhằm mục đích cung cấp cái nhìn rõ ràng về tình hình nghiên cứu hiện tại liên quan đến việc sử dụng SVM trong dự đoán lỗi phần mềm, từ đó định hình nền tảng lý thuyết và thực tiễn cho đề tài nghiên cứu.**

## **2.3. Tên đề tài, tính thời sự và tầm quan trọng của đề tài**

**2.3.1 Tên đề tài**

Đề tài: “Phân lớp dữ liệu dự đoán lỗi phần mềm dựa trên thuật toán SVM.”

**2.3.2 Tính thời sự và tầm quan trọng của đề tài**

* **Tính thời sự**:
* Trong bối cảnh phát triển mạnh mẽ của ngành công nghệ thông tin, việc đảm bảo chất lượng và độ tin cậy của phần mềm trở thành một yêu cầu không thể thiếu. Thuật toán SVM, với khả năng phân lớp dữ liệu một cách chính xác, đang ngày càng trở nên quan trọng trong việc dự đoán và phát hiện lỗi phần mềm.
* Sự phức tạp của các hệ thống phần mềm ngày nay, cùng với yêu cầu về thời gian phát triển nhanh và chi phí thấp, đòi hỏi phải có những phương pháp tiên tiến trong việc phân tích và quản lý lỗi. SVM cung cấp một phương pháp hiệu quả để đáp ứng những yêu cầu này.
* Việc áp dụng AI và machine learning trong quản lý và phát triển phần mềm đang mở ra những cơ hội mới. SVM, như một công cụ mạnh mẽ trong lĩnh vực này, đóng một vai trò không thể thiếu trong việc cải thiện quá trình này.
* **Tầm quan trọng**:
* Việc nắm vững và áp dụng thuật toán SVM trong dự đoán lỗi phần mềm không chỉ giúp tăng cường bảo mật và độ tin cậy của các sản phẩm phần mềm mà còn góp phần vào việc giảm thiểu chi phí và thời gian phát triển.
* Thuật toán SVM có khả năng tận dụng dữ liệu lịch sử và hiện tại để dự đoán lỗi, từ đó hỗ trợ các nhà phát triển trong việc tập trung nguồn lực vào những phần quan trọng nhất của quá trình kiểm thử.
* Trong lĩnh vực công nghệ thông tin và phát triển phần mềm, kiến thức và kỹ năng ứng dụng SVM trong dự đoán lỗi là vô cùng quan trọng, không chỉ đối với nhà phát triển phần mềm và các bên liên quan khác.
* Đề tài này không chỉ đáp ứng nhu cầu thực tiễn của ngành công nghệ thông tin trong việc phát triển phần mềm mà còn phản ánh xu hướng tiên tiến trong việc áp dụng các thuật toán học máy để nâng cao chất lượng và hiệu quả công việc.

# **Phần 3: Nội dung và phương pháp nghiên cứu**

## **3.1. Địa điểm và thời gian thực tập**

Địa điểm: Công Ty TNHH Hana Korea Việt Nam

Thời gian thực tập: từ 15/01/2024 đến 07/04/2024

## **3.2. Nội dung nghiên cứu**

Đề tài “Phân lớp dữ liệu dự đoán lỗi phần mềm dựa trên thuật toán SVM” là một chủ đề quan trọng trong lĩnh vực phát triển phần mềm và học máy. Dưới đây là một số nội dung nghiên cứu cụ thể về đề tài này:

* Tìm hiểu lý thuyết phân lớp dữ liệu và thuật toán SVM:
* Giới thiệu về học máy và phân lớp dữ liệu: Tìm hiểu cơ bản về học máy, sự phân biệt giữa học giám sát và không giám sát, và tầm quan trọng của việc phân lớp dữ liệu trong học máy.
* Thuật toán SVM (Máy Vector Hỗ Trợ): Lịch sử phát triển, nguyên lý hoạt động của SVM, cách thức xây dựng mặt phân cách và ứng dụng của SVM trong việc phân lớp dữ liệu.
* Ứng dụng của SVM trong phân lớp lỗi phần mềm: Tìm hiểu cách thức áp dụng SVM để phân biệt và dự đoán các loại lỗi trong phát triển phần mềm, từ đó giúp nâng cao chất lượng và độ tin cậy của sản phẩm.
* Cài đặt SVM với Python:
* Tìm hiểu về ngôn ngữ Python và thư viện học máy: Giới thiệu về Python và các thư viện học máy phổ biến như scikit-learn, TensorFlow hoặc PyTorch, cùng với cách sử dụng chúng để triển khai SVM.
* Cài đặt mô hình SVM cho bài toán phân lớp lỗi phần mềm: Sử dụng Python và thư viện scikit-learn để xây dựng mô hình SVM, huấn luyện mô hình với dữ liệu về lỗi phần mềm, và thực hiện phân lớp để dự đoán lỗi.
  1. **. Phương pháp nghiên cứ****u:**
* Tìm hiểu và phân tích cơ sở lý thuyết của thuật toán SVM: Thực hiện thông qua tài liệu tham khảo, bài giảng, và các nguồn thông tin đáng tin cậy trên internet.
* Phân tích và đánh giá ứng dụng thực tế của SVM trong dự đoán lỗi phần mềm: Nghiên cứu các trường hợp sử dụng đã được công bố, bao gồm việc áp dụng SVM trong các dự án thực tế và kết quả đạt được.
* Triển khai thực nghiệm: Xây dựng mô hình SVM sử dụng dữ liệu thực tế hoặc mô phỏng để đánh giá hiệu quả của thuật toán trong việc phân lớp và dự đoán lỗi phần mềm.

Đề tài này nhằm mục đích không chỉ tìm hiểu sâu rộng về thuật toán SVM và ứng dụng của nó trong lĩnh vực phát triển phần mềm mà còn nhằm thực hiện các nghiên cứu thực nghiệm để đánh giá hiệu quả của SVM trong việc dự đoán và phòng ngừa lỗi phần mềm, từ đó góp phần vào việc nâng cao chất lượng sản phẩm phần mềm.

# **Phần 4: Kế hoạch và dự kiến kết quả đạt được**

|  |  |  |  |
| --- | --- | --- | --- |
| **STT** | **Nội dung nghiên cứu** | **Ngày thực hiện** | **Kết quả cần đạt** |
| 1 | Tìm hiểu về học máy và thuật toán SVM. | 15/01/2024 đến 01/02/2024 | Báo cáo về học máy và thuật toán SVM |
| 2 | Tìm hiểu cách SVM được ứng dụng trong phân lớp dữ liệu và dự đoán lỗi phần mềm. | 01/02/2024 đến 15/02/2024 | Báo cáo về ứng dụng SVM trong phân lớp và dự đoán lỗi |
| 3 | Tìm hiểu về ngôn ngữ Python và thư viện học máy phổ biến. | 15/02/2024 đến 29/02/2024 | Báo cáo về Python và thư viện học máy |
| 4 | Xác định và phân tích dữ liệu thực nghiệm. | 01/03/2024 đến 03/03/2024 | Tài liệu đặc tả dữ liệu thực nghiệm |
| 5 | Thiết kế và xây dựng mô hình SVM. | 03/03/2024 đến 05/03/2024 | Mô hình SVM được thiết kế và xây dựng |
| 6 | Huấn luyện, tinh chỉnh mô hình SVM và đánh giá hiệu suất. | 05/03/2024 đến 31/03/2024 | Báo cáo về kết quả huấn luyện và đánh giá mô hình |
| 7 | Kiểm thử mô hình với dữ liệu thử nghiệm mới. | 01/04/2024 đến 05/04/2024 | Báo cáo về hiệu suất mô hình trên dữ liệu mới |
| 8 | Viết báo cáo | 06/04/2024 | Tài liệu báo cáo |
|  |  |  |  |

# **Phần 5: Tài liệu tham khảo**

1. SVM – Machine Learning cơ bản: <https://machinelearningcoban.com/2017/04/09/smv/>
2. Tài liệu lý thuyết: từ thầy cô hướng dẫn đồ án.