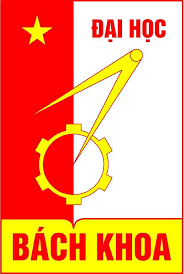
**TRƯỜNG ĐẠI HỌC BÁCH KHOA HÀ NỘI**

**VIỆN CÔNG NGHỆ THÔNG TIN VÀ TRUYỀN THÔNG**



**ĐỒ ÁN TỐT NGHIỆP**

**Xây dựng hệ thống xem phim online tích hợp hệ gợi ý bằng PredictionIO**

**NGUYỄN DANH NAM**

nam.nd166477@sis.hust.edu.vn

**Cử Nhân Công Nghệ - Công Nghệ Thông Tin**

|  |  |
| --- | --- |
| **Giảng viên hướng dẫn:** | TS. Nguyễn Nhật Quang |
| **Bộ môn:** | Hệ thống thông tin |
| **Viện:** | Công nghệ thông tin và truyền thông |
| **HÀ NỘI, 6/2020** | |

**PHIẾU GIAO NHIỆM VỤ ĐỒ ÁN TỐT NGHIỆP**

1. Thông tin về sinh viên
2. Mục đích nội dung của ĐATN
3. Các nhiệm vụ cụ thể của ĐATN
4. Lời cam đoan của sinh viên

Hà Nội, ngày 28 tháng 05 năm 2020

Tác giả ĐATN

Nguyễn Danh Nam

1. Xác nhận của giáo viên hướng dẫn về mức độ hoàn thành của ĐATN và cho phép bảo vệ

Hà Nội, ngày 21 tháng 05 năm 2019

Giáo viên hướng dẫn

TS. Nguyễn Nhật Quang

**Lời cảm ơn**

Trước khi vào nội dung chi tiết, trước hết em xin gửi lời cảm ơn chân thành tới thầy TS.Nguyễn Nhật Quang đã đồng hành cùng em trong kì học cuối cùng vừa qua, cảm ơn thầy vì sự hướng dẫn nhiệt tình, những lời khuyên hữu ích, những lời nhận xét chi tiết và tỉ mỉ.

Bên cạnh đó không thể thiếu được sự quan tâm, giúp đỡ của gia đình, bạn bè, đồng nghiệp đã hỗ trợ em trong quá trình thực hiện đồ án. Hơn nữa, trải qua 4 năm sinh viên tại Bách Khoa đầy khó khăn không thể thiếu được sự quan tâm, chỉ dạy nhiệt tình của các thầy cô, những người bạn, sự quan tâm của nhà trường. Xin gửi lời cảm ơn tới tất cả vì đã giúp em có được như ngày hôm nay.

**Tóm tắt đồ án**

Đồ án tốt nghiệp *Xây dựng hệ thống xem phim online tích hợp hệ ý PredictionIO* tập trung vào thiết kế chi tiết hệ thống và cài đặt, lập trình hệ thống và tìm hiểu kiến thức cơ bản về 1 số phương thức cơ bản trong hệ gợi ý.

Áp dụng các kiến thức đã học về khoa học máy tính, phân tích thiết kế hệ thống thông tin, công nghệ phần mềm… để viết tài liệu thiết kế phần mềm, thiết kế hệ thống. Tiếp đến là sử dụng các ngôn ngữ lập trình, thư viện, công cụ, dịch vụ bên ngoài để hỗ trợ cho việc cài đặt, lập trình tạo ra các thành phần cần thiết cho hệ thống.

Sinh viên thực hiện

Ký và ghi rõ họ tên

**MỤC LỤC**

[DANH MỤC HÌNH VẼ 5](#_Toc41587037)

[CHƯƠNG 1. GIỚI THIỆU 8](#_Toc41587038)

[1.1 Đặt vấn đề 8](#_Toc41587039)

[1.2 Mục đích và phạm vi đề tài 8](#_Toc41587040)

[1.3 Định hướng và giải pháp 8](#_Toc41587041)

[1.4 Cấu trúc tài liệu 8](#_Toc41587042)

[CHƯƠNG 2. CƠ SỞ LÝ THUYẾT 9](#_Toc41587043)

[CHƯƠNG 3. PHÂN TÍCH VÀ THIẾT KẾ HỆ THỐNG 10](#_Toc41587044)

[3.1 Phân tích yêu cầu 10](#_Toc41587045)

[**3.1.1 Yêu cầu chức năng 10**](#_Toc41587046)

[**3.1.2 Yêu cầu phi chức năng 10**](#_Toc41587047)

[3.2 Phân tích và thiết kế phần mềm 10](#_Toc41587048)

[**3.2.1 Phân tích cấu trúc và hành vi của các lớp tham gia ca sử dụng 10**](#_Toc41587049)

[**3.2.2 Phân tích sự tương tác của các lớp tham gia ca sử dụng 10**](#_Toc41587050)

[**3.2.3 Kiến trúc tổng thể của hệ thống 10**](#_Toc41587051)

[**3.2.4 Thiết kế chi tiết lớp 10**](#_Toc41587052)

[**3.2.5 Thiết kế cơ sở dữ liệu 10**](#_Toc41587053)

[**3.2.6 Thiết kế giao diện sử dụng hệ thống 10**](#_Toc41587054)

[CHƯƠNG 4. KẾT LUẬN 11](#_Toc41587055)

[TÀI LIỆU THAM KHẢO 12](#_Toc41587056)

[PHỤ LỤC 13](#_Toc41587057)

# **DANH MỤC HÌNH VẼ**

# **DANH MỤC BẢNG BIỂU**

**DANH MỤC CÁC TỪ VIẾT TẮT**

# CHƯƠNG 1. GIỚI THIỆU

## Đặt vấn đề

Trước khi internet phát triển như ngày nay, hầu hết hình thức xem phim của chúng ta đó là trên truyền hình vô tuyến hoặc rạp chiếu phim, tuy nhiên đôi lúc điều này khá bất tiện khi người xem phải phụ thuộc vào lịch chiếu của nhà đài.

Trong những năm gần đây internet đã phát triển mạnh mẽ, kéo theo đó là các dịch vụ giải trí cũng được phát triển theo để phục vụ con người. Đặc biệt phải kể đến các hình thức giải trí online như xem phim, xem liveshow, ca nhạc được rất ưa chuộng vì tính tiện lợi của chúng mang lại. Ở nhu cầu xem phim online thì phải kể đến các nhà cung cấp dịch vụ này nổi tiếng như Netflix, HBO, Amazon Prime Video…, tại Việt Nam cũng có một số dịch vụ xem phim online như Fim+, 123phim, FPT Play và hàng trăm các trang web cho miễn phí khác.

Chúng ta đang sống trong thời đại gọi là thời đại 4.0 mà ai cũng có thể ít nhất một lần nghe đến các cụm từ như trí tuệ nhân tạo, máy tự học, robot… các công nghệ này đang rất phát phiển và ứng dụng mạnh mẽ ở nhiều lĩnh vực, có thể kể đến như trong thương mại điện tử, từng cái nhấp chuột hay lịch sử mua hàng của chúng ta đều được thu nhập và lưu trữ, những dữ liệu này có vai trò quan trọng trong việc gợi ý các sản phẩm mà người mua có thể thích. Hay một số dịch vụ quảng cáo như Google Ads và Facebooks Ads cũng vậy, chỉ vừa mới xem hoặc tìm kiếm một mẫu quần áo thì một lúc sau đã có những quảng cáo liên quan đến quần áo xuất hiện trên new feed của Facebook hoặc trên các trang web có dùng dịch vụ Google Ads.

Không chỉ về quảng cáo, thương mại điện tử, thì phim ảnh cũng vậy, có người dùng thích thể loại phim này, nhóm khách hàng khác lại có gu thưởng thức khác. Dựa vào sở thích, hành vi của người dùng, các dịch vụ xem phim không chỉ giúp giải quyết được nhu cầu giải trí mà còn hiểu được người dùng có thể thích phim gì, thể loại nào, điều này làm tăng trải nghiệm sử dụng

## Mục đích và phạm vi đề tài

Áp dụng các kiến thức đã học được để thiết kế và lập trình cho hệ thống xem phim nhằm :

* Cung cấp cho người dùng các bộ phim trong kho phim của hệ thống, các tính năng giúp quản trị viên quản trị người dùng, quản lí kho phim trên nền tảng web.
* Trau dồi kỹ năng lập trình.
* Tích hợp hệ gợi ý mã nguồn mở Apache PredictionIO vào hệ thống để gợi ý phim.

## Công cụ và công nghệ sử dụng

* Ngôn ngữ lập trình: Python, Javascript.
* Thư viện, framework: Angular 8, Flask.
* Cơ sở dữ liệu: MySQL.
* Các ứng dụng khác được tích hợp: AWS S3, PredictionIO

## 1.4 Cấu trúc tài liệu

Tài liệu này bao gồm 4 phần chính:

1. Tổng quan về đồ án.
2. Cơ sở lý thuyết: các kiến thức cơ bản về hệ gợi ý, giới thiệu về PredictionIO.
3. Chi tiết về phân tích, thiết kế hệ thống xem phim đã xây dựng.
4. Kết luận và hướng phát triển.

Ngoài ra các tài liệu tham khảo, phụ lục cũng được cung cấp chi tiết ở phần cuối của tài liệu.

# CHƯƠNG 2. CƠ SỞ LÝ THUYẾT

### 2.1 Khái niệm cơ bản về hệ gợi ý

Như đã giới thiệu qua về một vài ví dụ về ứng dụng của hệ gợi ý trong phần mở đầu ta cũng có thể hình dung phần nào về khái niệm của hệ gợi ý.

Hệ gợi ý hay hệ thống khuyến nghị có tên tiếng anh là Recommendation System hoặc Recommender System, có chức năng đưa ra các gợi ý, những gợi ý này là gì thì sẽ tuỳ theo bài toán cụ thể nó được áp dụng. Lấy một bài toán cụ thể tại một trang thương mại điện tử, hệ gợi ý thường để sử dụng để gợi ý sản phẩm, lấy ví dụ:

* Amazon gợi ý cho người mua các mặt hàng mà người dùng có thể thích hoặc những mặt hàng tương tự với những thứ mà người mua đang quan tâm.
* Netflix gợi ý các phim người dùng có thể thích, muốn xem.
* Facebook gợi ý kết bạn, Youtube đề xuất video cho người xem.

Có nhân tố chính trong hệ gợi ý đó là *user* (người dùng) và *item* (sản phẩm), hệ gợi ý tìm ra những *item* và gợi ý cho *user*. RS được chia thành hai nhóm:

1. Content-based system: gợi ý theo các đặc tính, đặc điểm của sản phẩm. Ví dụ như một người rất thích xem các bộ phim về hành động, bạo lực, kinh dị, hệ gợi ý sẽ sử dụng dữ kiện này tìm ra những bộ phim có thể loại tương tự.
2. Collaborative filtering: gợi ý theo sự tương tự giữa người dùng hoặc sản phẩm, những người dùng có sở thích tương tự nhau dựa vào các dữ liệu về mức độ quan tâm của người dùng cho các sản phẩm.

Vai trò của RS rất quan trọng trong việc tạo trải nghiệm người dùng tốt hơn, đặc biệt trong kinh doanh nó có thể làm tăng sự hài lòng của khách hàng, tăng lợi nhuận. Việc ứng dụng, tích hợp RS ngày nay cũng khá đơn giản, ngoài việc có thể tự xây dựng hệ gợi ý thì các ứng dụng mã nguồn mở, các dịch vụ như PredictionIO, Amazon Personalize,… giúp ta có thể tích hợp RS vào hệ thống một cách dễ dàng. Hệ thống xem phim online của đồ án đã được tích hợp hệ gợi ý PredictionIO vào để gợi ý phim cho người xem. Chi tiết về PredictionIO sẽ được trình bày ở dưới đây.

### 2.2 PredictionIO

PredictionIO có tên đầy đủ là Apache PredictionIO và có khái niệm là:

*“Apache PredictionIO® is an open source Machine Learning Server built on top of a state-of-the-art open source stack for developers and data scientists to create predictive engines for any machine learning task”.*

Theo định nghĩa trên của trang chủ [PredictionIO](https://predictionio.apache.org/), ta có thể hiểu đơn giản rằng đây là một ứng dụng mà nguồn mở, được xây dựng dựa trên các ứng dụng mã nguồn mở khác để xử lí các công việc trong học máy.

Thành phần của PredictionIO gồm ba phần hợp thành:

1. PredictionIO platform: là thành phần chứa “machine learing stack” mã nguồn mở để quản lí các Engines.
2. Event Server “open source machine learning analytics layer for unifying events from multiple platforms”: là thành phần chịu trách nhiệm việc thu nhận các dữ liệu.
3. Template Gallery: chứa các engine templates.

# 

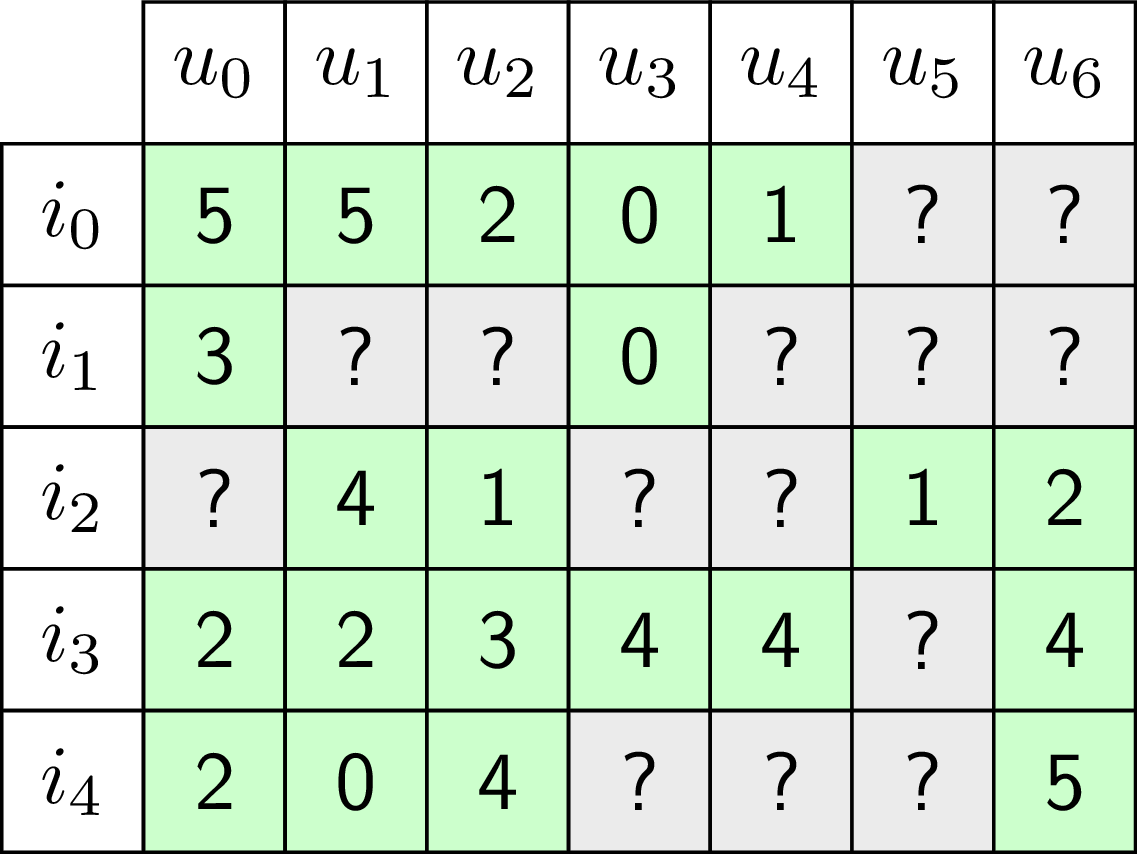
Hình 1 Mô hình minh hoạ cấu trúc của việc tích hợp PredictionIO vào hệ thống

Enginechính là một engine template đã được build và triển khai, lí do cho việc có chữ “template” là do các engine này là mã nguồn mở, được thiết kế để dễ dàng chỉnh sửa, tuỳ chỉnh. Thành phần này chịu trách nhiệm xử lí các công việc của quá trình tạo ra kết quả gợi ý. Dữ liệu dùng để training được lấy từ Event Server và tạo ra *model*. Engine sẽ được triển khai như là web service, nhờ đó kết quả gợi ý có thể lấy được thông qua REST API.

Trong một engine có thể có một hoặc nhiều thuật toán gợi ý, có engine thì có chức năng gợi ý những sản phẩm tương tự( Similar Product) có engine lại đa năng hơn khi thực thi rất nhiều thuật toán gợi ý như là Universal Recommender.

Trong hệ thống xem phim online này, hai engines được triển khai:

1. [Recommendation](https://predictionio.apache.org/templates/recommendation/quickstart/): dùng để tìm ra những bộ phim mà người xem có thể thích dựa vào kết quả đánh giá phim. Dữ liệu yêu cầu để training là các đánh giá của người dùng về các bộ phim.



Hình 2 Ma trận đánh giá

1. [Similar Product](https://predictionio.apache.org/templates/similarproduct/quickstart/): tìm ra các bộ phim tương tự với phim mà người dùng đang xem dựa trên các dữ liệu về thể loại của phim, lịch sử xem phim của người dùng.

Hai engine đều sử dụng phương pháp gợi ý Matrix factorization-based collaborative filtering bằng thư viện Spark MLlib về cơ bản cũng sẽ dựa trên các đặc tính của user, item ( nếu có) và mức độ quan tâm cuả user cho item để đưa ra các kết quả dự đoán.

# CHƯƠNG 3. PHÂN TÍCH VÀ THIẾT KẾ HỆ THỐNG

## **Phân tích yêu cầu**

Cung cấp mô tả chi tiết về các yêu cầu cho hệ thống xem phim trực tuyến. Bao gồm các ràng buộc và các yêu cầu phần mềm của hệ thống được xây đựng.

### **Yêu cầu chức năng**

* Các tác nhân trong hệ thống:
* Admin: quản trị viên hệ thống.
* User: người dùng (có tài khoản) của hệ thống.
* Guest: khách vãng lai (chưa đăng nhập).
* Sơ đồ usecase tổng quan:

**A close up of a map

Description automatically generated**

* Đặc tả usecase Đăng ký tài khoản

|  |  |
| --- | --- |
| Mã usecase | UC1 |
| Tên usecase | Đăng ký tài khoản |
| Tác nhân | Guest |
| Mô tả | Khách vãng lai tạo tài khoản dùng để đăng nhập vào hệ thống. |
| Sự kiện kích hoạt | Người dùng bấm vào nút “Đăng ký” trên giao diện trang chủ hoặc trên form đăng nhập. |
| Điều kiện tiên quyết | Không có |
| Luồng xử lí chính | * + 1. Người dùng click vào “Đăng ký”.     2. Giao diện hiển thị form đăng ký.     3. Người dùng điền các thông tin: tên đăng nhập (username), email, mật khẩu của người dùng trên form đăng ký.     4. Hệ thống kiểm tra thông tin và lưu dữ liệu đăng ký vào cơ sở dữ liệu. Hiển thị thông báo việc đăng ký tài khoản.     5. Gửi mail tới địa chỉ email của người dùng để xác thực tài khoản. |
| Luồng xử lí ngoại lệ | 4.a Thông tin không hợp lệ: thông báo lại cho người dùng. |

* Đặc tả usecase Đăng nhập
* Đặc tả usecase Đăng xuất
* Đặc tả usecase Lấy lại mật khẩu
* Đặc tả usecase Chỉnh sửa thông tin tài khoản
* Phân rã usecase Khám phá kho phim
* Đặc tả usecase Tìm kiếm
* Đặc tả usecase Xem thông tin phim
* Đặc tả usecase Xem trực tuyến
* Đặc tả usecase Đánh giá phim
* Đặc tả usecase Bình luận
* Đặc tả usecase Thêm phim vào danh sách yêu thích
* Đặc tả usecase Xoá phim khỏi danh sách yêu thích
* Đặc tả usecase Chia sẻ phim
* Đặc tả usecase Nhận gợi ý phim
* Phân rã usecase Quản lí Phim
* Đặc tả usecase Tìm kiếm phim
* Đặc tả usecase Thêm phim
* Đặc tả usecase Xem chi tiết phim
* Đặc tả usecase Sửa phim
* Đặc tả usecase Xoá phim
* Phân rã usecase Quản lí thể loại phim
* Đặc tả usecase Tìm kiếm thể loại phim
* Đặc tả usecase Thêm thể loại
* Đặc tả usecase Sửa thể loại phim
* Đặc tả usecase Xem thông tin thể loại
* Đặc tả usecase Xoá thể loại
* Phân rã usecase Quản lí người dùng
* Đặc tả usecase Tìm kiếm người dùng
* Đặc tả usecase Xem thông tin người dùng
* Đặc tả usecase Chỉnh sửa quyền của tài khoản
* Đặc tả usecase Khoá người dùng
* Đặc tả usecase Bỏ khoá người dùng
* Đặc tả usecase Xem thống kê
* Phân rã usecase Quản lí bình luận
* Đặc tả usecase Xem danh sách bình luận
* Đặc tả usecase Xoá bình luận

### Yêu cầu phi chức năng

Không có yêu cầu phi chức năng cho hệ thống này.

## 3.2 Phân tích và thiết kế phần mềm

Các thông tin thiết kế bao gồm: phân tích các lớp tham gia ca sử dụng, phân tích sự tương tác, thiết kế kiến trúc tổng thể của hệ thống, thiết kế chi tiết các lớp, thiết kế giao diện của ứng dụng và thiết kế cơ sử dữ liệu.

### 3.2.1 Phân tích cấu trúc và hành vi của các lớp tham gia ca sử dụng

* Đăng ký tài khoản
* Đăng nhập
* Đăng xuất
* Lấy lại mật khẩu
* Chỉnh sửa thông tin tài khoản
* Khám phá kho phim
* Tìm kiếm
* Xem thông tin phim
* Xem trực tuyến
* Đánh giá phim
* Bình luận
* Thêm phim vào danh sách yêu thích
* Xoá phim khỏi danh sách yêu thích
* Chia sẻ phim
* Nhận gợi ý phim
* Quản lí Phim
* Tìm kiếm phim
* Thêm phim
* Xem chi tiết phim
* Sửa phim
* Xoá phim
* Quản lí thể loại phim
* Tìm kiếm thể loại phim
* Thêm thể loại
* Sửa thể loại phim
* Xem thông tin thể loại
* Xoá thể loại
* Quản lí người dùng
* Tìm kiếm người dùng
* Xem thông tin người dùng
* Chỉnh sửa quyền của tài khoản
* Khoá người dùng
* Bỏ khoá người dùng
* Xem thống kê
* Quản lí bình luận
* Xem danh sách bình luận
* Xoá bình luận

### 3.2.2 Phân tích sự tương tác của các lớp tham gia ca sử dụng

* Đăng ký tài khoản
* Đăng nhập
* Đăng xuất
* Lấy lại mật khẩu
* Chỉnh sửa thông tin tài khoản
* Khám phá kho phim
* Tìm kiếm
* Xem thông tin phim
* Xem trực tuyến
* Đánh giá phim
* Bình luận
* Thêm phim vào danh sách yêu thích
* Xoá phim khỏi danh sách yêu thích
* Chia sẻ phim
* Nhận gợi ý phim
* Quản lí Phim
* Tìm kiếm phim
* Thêm phim
* Xem chi tiết phim
* Sửa phim
* Xoá phim
* Quản lí thể loại phim
* Tìm kiếm thể loại phim
* Thêm thể loại
* Sửa thể loại phim
* Xem thông tin thể loại
* Xoá thể loại
* Quản lí người dùng
* Tìm kiếm người dùng
* Xem thông tin người dùng
* Chỉnh sửa quyền của tài khoản
* Khoá người dùng
* Bỏ khoá người dùng
* Xem thống kê
* Quản lí bình luận
* Xem danh sách bình luận
* Xoá bình luận

### 3.2.3 Kiến trúc tổng thể của hệ thống

* Mô hình kiến trúc hoạt động của hệ thống
* Sơ đồ triển khai hệ thống

### 3.2.4 Thiết kế chi tiết lớp

* Gói View
* Gói Controller
* Gói Model

### 3.2.5 Thiết kế cơ sở dữ liệu

* Biểu đồ quan hệ thực thể dữ liệu (Entity Relationship Diagram)
* Thiết kế chi tiết các bảng dữ liệu

### 3.2.6 Thiết kế giao diện sử dụng hệ thống

# CHƯƠNG 4. KẾT LUẬN

# TÀI LIỆU THAM KHẢO

# PHỤ LỤC