

**TỔNG CÔNG TY VIỄN THÔNG VIETTEL**

**KHỐI CÔNG NGHỆ THÔNG TIN**

**PYC THỬ NGHIỆM ĐÁNH GIÁ CÁC MÔ HÌNH TREE-BASED VÀ RANKING ALGORITHM**

**TÀI LIỆU THIẾT KẾ CHI TIẾT**

**Mã hiệu dự án:** **MyViettel\_0008**

**Mã hiệu tài liệu: MyViettel\_0008**

**<Địa điểm, Thời gian>**

**BẢNG GHI NHẬN TIẾN ĐỘ**

\*A - Tạo mới, M - Sửa đổi, D - Xóa bỏ

| **Ngày**  **bàn giao** | **Yêu cầu** | **A\***  **M, D** | **Trạng thái** | **Đầu mối KH** | **Mô tả** | **Ghi chú** |
| --- | --- | --- | --- | --- | --- | --- |
| 20/4/2023 | Đánh giá các nhóm mô hình Machine learning và Deep learning | A | Hoàn thành | PTC, PKH | * Mô tả chi tiết, đánh giá performance mô hình |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |

**MỤC LỤC**

[**I. Mô tả bộ dữ liệu 9**](#_heading=h.ay2vm15aknsu)

[1.1. Bộ dữ liệu MovieLens-20M 9](#_heading=h.oyo1y9mnb0qm)

[**II. Build mô hình H+VAMP 10**](#_heading=h.lyny3gviq1n7)

[2.1 Cài đặt môi trường 10](#_heading=h.z002uvvff06p)

[2.1.1. Tạo môi trường ảo 10](#_heading=h.ib3v7dit3dzb)

[2.1.2. Set up các thư viện 10](#_heading=h.xz9q255l6lqj)

[2.2 Xử lý tập dữ liệu 10](#_heading=h.at78sl80yv5q)

[2.2.1 Nối các tập dữ liệu log và film 10](#_heading=h.yacfd3si25sk)

[2.2.2. Loại bỏ các giá trị invalid 10](#_heading=h.avt3wbhrda5z)

[2.2.3. Định dạng lại các trường dữ liệu 10](#_heading=h.90tl6xtl04l2)

[2.2.4. Kết nối log data và film data 10](#_heading=h.z4gfzirulxbc)

[2.2.5. Loại bỏ các trường dữ liệu không cần thiết 10](#_heading=h.esx24jnp5hc7)

[2.2.6. Loại bỏ các user ít hoạt động và outlier 10](#_heading=h.ytxpy2jruwol)

[2.2.3 Chia tách tập dữ liệu 10](#_heading=h.oyp6jh1oksuj)

[2.2.4. Lưu tập dữ liệu 10](#_heading=h.6cq658pheqbn)

[2.3 Quá trình luyện mô hình 10](#_heading=h.n7pca4rfd60w)

[2.3.1 Khởi tạo tham số 10](#_heading=h.pqgevca96dor)

[2.3.2 Quá trình huấn luyện: 11](#_heading=h.nom5i13ehc9y)

[2.3.3. Xây dựng hàm đánh giá 11](#_heading=h.wtlyzm33a7a9)

[**III. Fine-Tune và đánh giá performance mô hình HVamp 11**](#_heading=h.62m72o386zno)

[3.1 Fine Tune mô hình SetRank 11](#_heading=h.19h9gp6v0prq)

[3.1.1. Thông số hyperparameter: 11](#_heading=h.y8amfdfsdjo3)

[3.1.2. Kết quả train mô hình: 11](#_heading=h.bi1miqi8yuia)

[3.1.3. Kết quả test: 11](#_heading=h.5xvtvts2ykbd)

[3.2 Fine Tune mô hình BM25 11](#_heading=h.ptd1utb1gtze)

[3.2.1. Thông số hyperparameter: 11](#_heading=h.8qz4aodq0ere)

[3.2.2. Kết quả train mô hình: 11](#_heading=h.jt124ox4qu9e)

[3.2.3. Kết quả test: 11](#_heading=h.xyfwskb536tu)

[3.3 Fine Tune mô hình LambdaMART 11](#_heading=h.wvndk85645j7)

[3.3.1. Thông số hyperparameter: 11](#_heading=h.trfe6qva44s7)

[3.3.2. Kết quả train mô hình: 11](#_heading=h.pidxw3sh8hoa)

[3.3.3 Kết quả test: 11](#_heading=h.p1h5p24rnkds)

[3.4 Fine Tune mô hình 11](#_heading=h.hw06f5jwt3e4)

[3.4.1. Thông số hyperparameter: 11](#_heading=h.eecgspzctyqs)

[3.4.2. Kết quả train mô hình: 11](#_heading=h.rhimskw6mdwl)

[3.4.3. Kết quả test: 11](#_heading=h.5wmk3g9lsmkw)

[3.5 Fine Tune mô hình 11](#_heading=h.9plwz5anq11m)

[3.5.1. Thông số hyperparameter: 11](#_heading=h.3f3wgp5ale9b)

[3.5.2. Kết quả train mô hình: 11](#_heading=h.kybl42sp3wwt)

[3.5.3. Kết quả test: 11](#_heading=h.jubrfzze9zc5)

[3.6 Fine Tune mô hình 11](#_heading=h.usn6isj9np7h)

[3.6.1. Thông số hyperparameter: 11](#_heading=h.87w9qos4hbg6)

[3.6.2. Kết quả train mô hình: 11](#_heading=h.l82gn3y86e5l)

[3.6.3. Kết quả test: 11](#_heading=h.u9whnsssx1y)

[3.7 Fine Tune mô hình lần 7 11](#_heading=h.gszhu0o0i7bt)

[3.7.1. Thông số hyperparameter: 11](#_heading=h.fjq5jg1wuey2)

[3.7.2. Kết quả train mô hình: 12](#_heading=h.muguj6juyh5r)

[3.7.3. Kết quả test: 14](#_heading=h.49j45p82fpu9)

[3.8 Fine Tune mô hình lần 8 14](#_heading=h.ivgczsgdtgwm)

[3.8.1. Thông số hyperparameter: 14](#_heading=h.awxji053ytm4)

[3.8.2. Kết quả train mô hình: 15](#_heading=h.72vtfala5aif)

[3.8.3. Kết quả test: 17](#_heading=h.oyygdi1i58vc)

[3.9 Fine Tune mô hình lần 9 17](#_heading=h.3o6qzz4q7rh1)

[3.9.1. Thông số hyperparameter: 17](#_heading=h.dpze7epfg9wu)

[3.9.2. Kết quả train mô hình: 18](#_heading=h.bbdzngcbzxfx)

[3.9.3. Kết quả test: 19](#_heading=h.vembx280e74n)

[3.10 Fine Tune mô hình lần 10 19](#_heading=h.byb9pstv44zi)

[3.10.1. Thông số hyperparameter: 19](#_heading=h.3i9lp6aeadvt)

[3.10.2. Kết quả train mô hình: 20](#_heading=h.r7p3xq1kddcu)

[3.10.3. Kết quả test: 22](#_heading=h.o9gkr8k6cll4)

[3.11 Fine Tune mô hình lần 11 23](#_heading=h.mu0b6311ty4k)

[3.11.1. Thông số hyperparameter: 23](#_heading=h.glhs5qk273kw)

[3.11.2. Kết quả train mô hình: 23](#_heading=h.exfoltbw95rz)

[3.11.3. Kết quả test: 26](#_heading=h.o3cx191p1z1)

[**IV. Build mô hình IGMC 26**](#_heading=h.dkb0ksn1q93b)

[4.1 Cài đặt môi trường 26](#_heading=h.a1z2ylrf6fnp)

[4.1.1. Tạo môi trường ảo 26](#_heading=h.79wjgrcqjjus)

[4.1.2. Set up các thư viện 26](#_heading=h.ljnq2uw05yh4)

[4.2 Tập dữ liệu 33](#_heading=h.2bldgfaxvf4m)

[4.2.1 Xử lý dữ liệu đầu vào 34](#_heading=h.yv04q6uy8bb)

[4.2.2 Xử lý đọc dữ liệu 34](#_heading=h.sfsb0p4s338x)

[4.2.3 Chia tập dữ liệu 37](#_heading=h.11fu4c6jfmju)

[4.2.4 Chuẩn hóa dữ liệu 42](#_heading=h.holzepfo3yoo)

[4.3 Huấn luyện mô hình 44](#_heading=h.zebqgy67qlvx)

[4.3.1 Khởi tạo tham số 44](#_heading=h.1q6lf1dnamfx)

[4.3.2 Quá trình huấn luyện: 44](#_heading=h.jjq1dbe7s6li)

[4.3.3 Đánh giá 52](#_heading=h.owsgh769ycd2)

[**V. Fine-Tune và đánh giá performance mô hình … 53**](#_heading=h.qc5pg9ob4mlv)

[5.1 Fine Tune mô hình lần 1 53](#_heading=h.bx7o8lmlamk6)

[5.1.1. Thông số hyperparameter: 53](#_heading=h.2dmrx0l7hltx)

[5.1.2. Kết quả train mô hình: 54](#_heading=h.lxl3zyzhlix5)

[5.1.3. Kết quả test: 57](#_heading=h.hxcfvkyrvt9l)

[5.2 Fine Tune mô hình lần 2 57](#_heading=h.vgb8rlrsx947)

[5.2.1. Thông số hyperparameter: 57](#_heading=h.g8tnxpdfw8c3)

[5.2.2. Kết quả train mô hình: 58](#_heading=h.oomoq2eseenl)

[5.2.3. Kết quả test: 58](#_heading=h.ql2njam3zvr0)

[5.3 Fine Tune mô hình lần 3 58](#_heading=h.n8cuucapd3r2)

[5.3.1. Thông số hyperparameter: 58](#_heading=h.ax2oqchn20r)

[5.3.2. Kết quả train mô hình: 59](#_heading=h.cvv9csj1ao0r)

[5.3.3 Kết quả test: 60](#_heading=h.g48i5jmpw590)

[5.4 Fine Tune mô hình lần 4 60](#_heading=h.2m3n5j350b7o)

[5.4.1. Thông số hyperparameter: 60](#_heading=h.n1z95t7yis13)

[5.4.2. Kết quả train mô hình: 61](#_heading=h.5h6cvjdqemay)

[5.4.3. Kết quả test: 63](#_heading=h.b9uwzxp7731n)

[5.5 Fine Tune mô hình lần 5 64](#_heading=h.7cdixj5mgt8u)

[5.5.1. Thông số hyperparameter: 64](#_heading=h.5pa0w74wddjg)

[5.5.2. Kết quả train mô hình: 64](#_heading=h.thtpwo4hj7hh)

[5.5.3. Kết quả test: 67](#_heading=h.gn9nd2exew3m)

[5.6 Fine Tune mô hình lần 6 67](#_heading=h.3d37ajqm5x69)

[5.6.1. Thông số hyperparameter: 67](#_heading=h.9bzn42d16iy)

[5.6.2. Kết quả train mô hình: 68](#_heading=h.uzhjbc2bn0c6)

[5.6.3. Kết quả test: 70](#_heading=h.q8rv2f3y2onw)

[5.7 Fine Tune mô hình lần 7 71](#_heading=h.crm1165lt3pl)

[5.7.1. Thông số hyperparameter: 71](#_heading=h.71e21css8mvt)

[5.7.2. Kết quả train mô hình: 71](#_heading=h.z3y316o1cbc6)

[5.7.3. Kết quả test: 73](#_heading=h.hvr7uxncxdhw)

**TRANG KÝ**

Người lập: <Ngày>

<Chức danh>

Người xem xét: <Ngày>

<Chức danh>

Người xem xét: <Ngày>

<Chức danh>

Người phê duyệt: <Ngày>

<Chức danh>

# 

# I. Xử lý dữ liệu

## 1.1 Sinh data fake

## 1.2 Tổng hợp feature dữ liệu

## 1.3 Xử lý Null

## 1.4 Xử lý outlier

## 1.5 Thực hiện feature engineering

## 1.6 Lựa chọn feature

## 1.7 Visualize dữ liệu

# II. Fine-Tune và test mô hình PRM

## 2.1 Training mô hình

## 2.2 Đánh giá kết quả

## 2.3 Fine Tune mô hình lần 1

### 2.3.1. Thông số hyperparameter:

### 2.3.2. Kết quả train mô hình:

### 2.3.3. Kết quả test:

## 2.4 Fine Tune mô hình lần 2

### 2.4.1. Thông số hyperparameter:

### 2.4.2. Kết quả train mô hình:

### 2.4.3. Kết quả test:

## 2.5 Fine Tune mô hình lần 3

### 2.5.1. Thông số hyperparameter:

### 2.5.2. Kết quả train mô hình:

### 2.5.3. Kết quả test:

# III. Fine-Tune và test mô hình DCN-v2

## 3.1 Training mô hình

## 3.2 Đánh giá kết quả

## 3.3 Fine Tune mô hình lần 1

### 3.3.1. Thông số hyperparameter:

### 3.3.2. Kết quả train mô hình:

### 3.3.3. Kết quả test:

## 3.4 Fine Tune mô hình lần 2

### 3.4.1. Thông số hyperparameter:

### 3.4.2. Kết quả train mô hình:

### 3.4.3. Kết quả test:

## 3.5 Fine Tune mô hình lần 3

### 3.5.1. Thông số hyperparameter:

### 3.5.2. Kết quả train mô hình:

### 3.5.3. Kết quả test:

## 3.6. Kết quả model cuối