

WEEKLY REPORT

Người thực hiện: Lê Ngô Yến Phương, Võ Minh Khôi, Trần Châu Thanh Tâm

Data advising

Instructors: Linh Nguyễn, Tạ Khánh Chi, Nguyễn Hữu Tình

Date: 16/1/2019

# Công việc của team trong tuần này

* Tối ưu thuật toán, hiệu suất cào post trên facebook
* Code cào post từ các trang web tuyển dụng
* Tìm hiểu kỹ thuật visualize dữ liệu
* Tối ưu tổng thể các phương pháp cào dữ liệu
* Phân tích code của các thành viên
* Phân chia công việc cho nhau trên trello

## Công việc cụ thể từng cá nhân

* Khôi:
* Tối ưu lại facebook
* Code glassdoor, vietnamworks
* Fix bug, merge code của các thành viên
* Phương:
* Code itviec, jobstreet, indeed, careerbuilder
* Tối ưu code, fix bug
* Tổng hợp dữ liệu, report
* Tâm:
  + Tối ưu code facebook
  + Merge code
  + Visualize dữ liệu, report

**TỔNG HỢP DỮ LIỆU**

1. Dữ liệu hiện tại được chủ yếu lấy từ 4 trang web: Itviec, careerbuilder, jobstreet và indeed. Ngoài ra Team A còn sử dụng data trên FB, tuy nhiên việc lọc dữ liệu còn tốn nhiều thời gian, nên chưa khai thác được tối đa tiềm năng của FB.
2. Về công việc, theo Team A nhận thấy, phần lớn công việc về Data Science là Data Engineer (83), Data Scientist (55), Data Analyst (243). (mục Data-related jobs).
3. Team A chúng em cào theo tag Data nên sẽ có những job có chữ data nhưng không phải là job về DS. Chúng em sẽ khắc phục sớm nhất có thể. Các thông tin phía dưới sẽ là thông tin theo tag data.
4. Về phân bố, tuyển dụng của các công ty tập trung nhiều ở Tp.HCM (75%), Hà Nội (17%), các thành phố khác chiếm tỷ lệ 8%. (mục Jobs locations).
5. Trong điều kiện ứng tuyển, yêu cầu về số năm kinh nghiệm là không thể bỏ qua. Số năm có sự giao động từ fresher đến 5 năm, trong đó 2 năm chiếm tỷ lệ cao nhất với 33%, tiếp theo là 3 năm với 24% và 1 năm với 23%, 5 năm với 16% còn lại là 4 năm chiếm chỉ 4%. Như vậy, ta nhận thấy thị trường việc làm tập trung ở 2-3 năm kinh nghiệm là chủ yếu. Ngoài ra thì khả năng kiếm việc làm của người mới bước vào lĩnh vực DS cũng khá khả quan với mức tuyển dụng kinh nghiệm 1 năm là 23%. (Mục Experience)
6. Về FB, Team A nhận thấy thời gian post trong các group tăng vào khoảng các tháng sau tết 3-4-5. Đây là thời gian các công ty tìm kiếm nhân lực.
7. Giữa các group FB, Team nhận thấy trong gần 2 năm trở lại đây, các group có sự phát triển ổn định, các post được đăng đều, với 3905 post trong năm 2018, lượng post được trải khá đều trong các tháng. (Mục Yearly Facebook Posts và Monthly Facebook Posts)
8. Giữa các trang web, trang indeed có lượng jobs về các ngành có tag data khá lớn so với các trang còn lại. Lý do là vì indeed hoạt động theo phương thức: đi lấy thông tin nghề nghiệp từ nhiều trang khác và đăng tải lại.

**Cảm nghĩ của team**

Sau một tuần làm việc thì Team A bọn em xin phép viết một bản report dở ẹc cho mọi người đọc/xem:

Bọn em đã gặp rất nhiều khó khăn trong việc vác cuốc đi cày bừa trên các trang web tìm việc làm. Có trang thì phải click *see more* để load thêm, có trang thì *next page*, có trang thì giấu bớt thông tin bắt mình phải đăng nhập tài khoản, có trang thì có đến 2 trang index khác nhau để redirect đến (một cách random) sau khi log in, đã vậy còn bị chặn ip do load quá nhiều, quá nhanh…  
Dù cây cuốc khá nặng nhưng mỗi lần cào được một đoạn ruộng thì bọn em lại quen với việc cày bừa hơn 1 chút, mệt hơn 1 chút, cào nhanh hơn 1 chút. Sau khi làm vài thửa ruộng thì bắt đầu thoải mái hơn, gặp cục đá chặn đường thì giải quyết cục đá nhanh hơn, nhưng gặp tảng đá to quá thì đành đi vòng qua…  
Cả team đều có sự cố gắng hơn trong tuần trước, cố gắng hoàn thành các task đã được chia, kết quả vẫn chưa thật sự đủ, càng tìm hiểu càng thấy nhiều hướng đi, nhiều cái mới muốn tìm hiểu thêm thêm và đào sâu vào nhưng thời gian là có hạn.

Các thành viên trong team dành rất nhiều thời gian để chỉ dẫn nhau nên đã thân thiết với nhau hơn. Gần như mỗi ngày trong team đều có report nhanh trong ngày để nhanh chóng cùng nhau giải quyết các vấn đề khó.

**Thống kê sơ bộ về dữ liệu cào được**

**DATA-RELATED JOBS**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Website** | **Data Engineer** | **Data Scientist** | **Data Analyst** | **Others** |
| **ITviec** | 11 | 4 | 3 | 24 |
| **jobstreet** | 33 | 17 | 43 | 381 |
| **indeed** | 30 | 28 | 144 | 408 |
| **Careerbuilder** | 9 | 6 | 53 | 175 |
| **Total** | 83 | 55 | 243 | 988 |

**JOBS LOCATIONS**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Websites** | **Hồ Chí Minh** | **Hà Nội** | **Bình Dương** | **Đồng Nai** | **Bà Rịa - Vũng Tàu** | **Hải Phòng** | **Cần Thơ** | **Kon Tum** | **Thái Bình** | **Long Xuyên** | **Bắc Ninh** | **Hưng Yên** | **Ðà Nẵng** | **Long An** |
| **ITviec** | 33 | 9 |  |  |  |  |  |  |  |  |  |  |  |  |
| **jobstreet** | 76 | 67 | 1 |  | 3 |  |  |  |  |  |  |  |  | 1 |
| **indeed** | 437 | 62 | 27 | 11 | 4 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |  |
| **Careerbuilder** | 208 | 29 | 12 | 7 | 3 |  |  |  |  |  |  |  |  | 1 |
| **Total** | 754 | 167 | 40 | 18 | 10 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 |

|  |  |
| --- | --- |
| Experience |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Kinh Nghiệm** | **IT Viec** | **Careerbuilder** | **Indeed** | **Job\_Street** | **Total** |
| **1 năm** | 5 | 31 | 59 | 78 | 173 |
| **2 năm** | 12 | 25 | 117 | 96 | 250 |
| **3 năm** | 5 | 34 | 93 | 53 | 185 |
| **4 năm** |  | 4 | 14 | 11 | 29 |
| **5 năm** | 1 | 10 | 53 | 58 | 122 |

**YEARLY FACEBOOK POSTS**

|  |  |  |  |
| --- | --- | --- | --- |
| **Group** | **2019** | **2018** | **2017** |
| **DATA SCIENCE AND BIG DATA** | 83 | 1416 |  |
| **ML-AP** | 1 | 81 | 71 |
| **VIETNAM DATA SCIENCE** | 8 | 399 | 259 |
| **IT JOB** | 44 | 2009 | 202 |
| **Sum** | 136 | 3905 | 532 |

**MONTHLY FACEBOOK POSTS**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Group** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** |
| **DATA SCIENCE AND BIG DATA** | 57 | 31 | 75 | 102 | 72 | 65 | 95 | 126 | 137 | 191 | 230 | 235 |
| **ML-AP** | 2 | 4 | 2 | 2 | 2 | 13 | 10 | 12 | 11 | 7 | 12 | 4 |
| **VIETNAM DATA SCIENCE** | 31 | 29 | 26 | 32 | 32 | 29 | 42 | 41 | 31 | 27 | 41 | 38 |
| **IT JOB** | 154 | 82 | 166 | 272 | 228 | 151 | 127 | 224 | 182 | 185 | 148 | 90 |
| **Sum** | 244 | 146 | 269 | 408 | 334 | 258 | 274 | 403 | 361 | 410 | 431 | 367 |