# **RESEARCH PLAN**

## **ML01 APACHE MAHOUT**

|  |
| --- |
| **Group name:** Team18  **List of members:**   * 18120228 Huỳnh Nhựt Quang * 18120253 Mai Ngọc Tú * 18120374 Nguyễn Minh Hiếu |

|  |
| --- |
| **Keywords:** Machine Learning, Recommender engines, Clustering, Classification, Frequent item-set mining  **Description:** Học máy đã phát triển được một khoảng thời gian tính tới hiên tại. Ngày nay khi mà công việc, dịch vụ gắn liền với Internet, đặc biệt là lĩnh vực kinh doanh như E**-**commerce, khiến dữ liệu phát sinh liên tục và sự ra đời của Bigdata. Vậy với lượng dữ liệu lớn như vậy, ta phải tìm cách áp dụng các thuật toán học máy nhằm tìm ra những thứ có ích. Đã có nhiều framework ra đời nhằm giải quyết vấn đề này và Mahout là một framwork tiêu biểu nhất.  **List of references:** 〈materials that have been used in this research〉   * Documentations, tutorials, technical reports, etc. * <http://mahout.apache.org/docs/latest/> * [Mahout Tutorial (tutorialspoint.com)](https://www.tutorialspoint.com/mahout/index.htm) * Source codes, pretrained models, demonstrative applications, etc. * Elipse * List of books, papers, or web pages * [edureka.co: Introduction to Mahout](https://www.edureka.co/blog/videos/introduction-to-mahout/) (10 Sep 2014) * Mahout in Action (October 4, 2011) * Apache Mahout Essentials (June 2015) |

|  |
| --- |
| **Presentation outline**  〈Complete the following sections with the content of your own topic〉   1. **Introduction:** an overview of the selected solution (i.e., a tool, a project or a product that provides great helps to Big Data research)  * Which problems could be solved by using this solution? * A brief history: motivations and evolution through years  1. **A deeper insight to the selected solution**    1. Major components and main functionalities: consider what are worth to be presented and support them with demonstrative figures and examples    2. Its applications in academic and/or industry activities    3. Its popularity (i.e., how many users choose this solution)    4. Identify other solutions that have similar functionalities and compare them on different aspects 2. **Demonstration**  * Anything that could be used to highlight the solution’s main functionalities, rather than being just an installation guide. * The more effort you pay (e.g., interesting story, significant coding, etc.), the higher credit you get.   Theo trang web chính của Apache Mahout:   * [ResearchGate](https://www.researchgate.net/), trang web uy tín cho khoa học và nghiên cứu dùng thuật toán của Mahout để tạo ra các gợi ý. * [Twitter](https://twitter.com/explore) sử dụng thuât toán LDA do Mahout cài đặt để tạo mô hình về sở thích của người dùng. * [Yahoo! Mail](https://mail.yahoo.com/?.intl=vn&.lang=vi-VN&.partner=none&.src=fp) ứng dụng khai phá dữ liệu trên Mahout.     Source: [enlyft.com](https://enlyft.com/tech/products/apache-mahout)   1. **Discussions and Conclusion** |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Weekly schedule:** 〈what each member has done every week〉   |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | |  | Week 01  (from  4/10 to 11/10) | Week 02  (from 11/10 to 18/10) | Week3  (from 18/10 to 25/10) | Week4  (from 25/10 to 1/11) | Week 5  (from  1/11 to 8/11) | Week 6  (from  8/11 to 15/11) | | 18120228 | Thảo luận kế hoạch | Tìm hiểu các ứng dụng thực tế | Nghiên cứu Clustering | Cài đặt thuật toán  Clustering |  | Kiểm tra và đánh giá | | 18120253 | Thảo luận kế hoạch | Chạy thử Tool Mahout | Nghiên cứu Classifi  cation | Cài đặt thuật toán  Classifi  cation |  | Kiểm tra và đánh giá | | 18120374 | Thảo luận kế hoạch | Chạy thử Tool Mahout | Nghiên cứu Re  commend | Cài đặt thuật toán  Re  commend |  | Kiểm tra và đánh giá | |

|  |
| --- |
| **Group self-evaluation**   * **Advantages:** * Làm quen với BigData, áp dụng mô hình học máy vào chúng. * Học thêm về Java code. * Biết thêm hệ sinh thái Apache. * **Disadvantages:** 〈any difficulties that you might encounter during research〉 |