

DECLARATION

I hereby declare that the Major Project-II work entitled “**A Recommender System for Online Social Network Using Hadoop in Large Scale User Generated Data**” which is being submitted to the Delhi Technological University, in partial fulfillment of requirements for the award of degree of Master of Technology (SWE) in the Department of Computer Science & Engineering, is a bonafide report of the Major Project-II carried out by me. The material contained in this report has not been submitted to any University or Institution for the award of any degree.

Swati Sharma
Roll no:- 2K13/SWE/22
M.Tech.(Software Engineering)
Email: swatisharma2311@gmail.com

CERTIFICATE

This is certify that the Major Project-II Report entitled “**A Recommender System for Online Social Network Using Hadoop in Large Scale User Generated Data**” is the work of Swati Sharma (Roll no. 2K13/SWE/22). This project was completed under my supervision and form a part of Master of Technology (Software Engineering) course curriculum in the Department of Computer Science & Engineering, Delhi Technological University, Delhi.

(**Mr. Manoj Sethi**)
Project Guide
Dept. of Computer Science & Engineering
Delhi Technological University

ACKNOWLEDGEMENT

First of all, I would like to express my deep sense of respect and gratitude to my project supervisor **Mr. Manoj Sethi** for providing the opportunity of carrying out this project and being the guiding force behind this work. I am deeply indebted to him for the support, advice and encouragement he provided without which the project could not have been a success.

Secondly, I am grateful to **Dr. O.P.Verma**, HOD, Computer Engineering Department, DTU for his immense support. I would also like to acknowledge Delhi Technological University library and staff for providing the right academic resources and environment for this work to be carried out.

Last but not the least I would like to express sincere gratitude to my parents and friends for constantly encouraging me during the completion of work.

Swati Sharma
University Roll no: 2K13/SWE/22
M.Tech (Software Engineering)
Department of Computer Science & Engineering
Delhi Technological University, Delhi – 110042

ABSTRACT

We are living in an age of Data and Information. Online social networks are contributing in enlargement of this data on high scale and Recommendation systems are helping industries to make this data useful for business purposes. It is helping to enhance the opportunities in online social data. Online social network generate large quantity of data from its users and recommendation system use this data for suggesting right piece of information to the user. But in the time of Big Data, processing large volumes of data generating suggestions is a difficult job. We are aiming to implement a combined approach for recommendation algorithm which include user-based collaborative filtering and item-based collaborative filtering using Apache Mahout, a machine learning tool, on Hadoop platform to reduced the time for recommendation generation and to provide a scalable system for processing large data sets efficiently.

Keywords: Recommendation system, Hadoop, Apache Mahout, Collaborative filtering.

CONTENTS

Declaration	II
Certificate	III
Acknowledgment	IV
Abstract	V
Contents	VI
List of Figures	VII
List of Tables	VIII
Chapter 1	1
INTRODUCTION	
1.1 An Overview	1
1.2 General Concept	2
1.3 Motivation	4
1.4 Related Work	6
1.5 Problem Statement	8
1.6 Organization of Thesis	9

Chapter 2 **10****RECOMMENDATION SYSTEM OVERVIEW**

2.1 Definition of Recommender System	10
2.2 Idea behind Recommender System	13
2.2.1 Information retrieval and filtering	13
2.3 Models for Recommender System	17
2.3.1 Collaborative filtering (CF) approach	18
2.3.2 Content-based filtering	19
2.3.3 Hybrid Model	19
2.4 Algorithms	20
2.4.1 Pearson correlation	20
2.4.2 Clustering algorithms	20
2.4.3 Other algorithms	22
2.5 Challenges with Recommender Systems	22
2.6 Applications of Recommendation System	24

Chapter 3 **26****RECOMMENDATION SYSTEM AND BIG DATA**

3.1 Big Data: large scale user generated data	26
3.2 Problem of scalability and prospective solutions	29
3.3 Apache Hadoop	29
3.4 Apache Mahout	32

Chapter 4	34
CASE STUDY ON RECOMMENDATION SYSTEM : AMAZON	
4.1 Item-to-Item Collaborative Filtering	35
4.2 How It Works	38
4.3 Scalability	39
Chapter 5	42
IMPLEMENTATION	
5.1 Proposed work	42
5.2 User-based CF	44
5.3 Item-based CF	45
5.4 Combined approach	46
5.5 System Configuration	47
Chapter 6	49
RESULTS AND ANALYSIS	
6.1 Result analysis	49
Chapter 7	54
CONCLUSION AND FUTURE WORK	
7.1 Conclusion	54
7.2 Future work	54
REFERENCES	56

LIST OF FIGURES

Figure 2.1: Information retrieval	14
Figure 2.2: Information filtering in RS	17
Figure 2.3: Models for RS	18
Figure 3.1: percentage of companies where big data improved the decision making	28
Figure 3.2: The base Hadoop framework.	31
Figure 4.1: “Your Recommendation” feature on amazon.com homepage	37
Figure 4.2: Amazon.com shopping cart recommendations	38
Figure 5.1: Combined approach for CF	44
Figure 5.1: User-based Collaborative Filtering	45
Figure 5.2: Item-based Collaborative Filtering	46
Figure 5.3: Combined approach for CF	47
Figure 6.1 : Speedup with respect to no. of nodes	51
Figure 6.2: Accuracy of recommendation	52
Figure 6.3: Speedup with respect to increase in dataset.	53

List of Tables

Table 1: System Configuration	48
-------------------------------	----