Developing a Scalable and Accurate Job Recommendation System with Distributed Cluster System using Machine Learning Algorithm

By

Timothy Dicky
11502006

BACHELOR'S DEGREE
in

Information Technology
Faculty of Engineering and Information Technology

SWISS GERMAN UNIVERSITY
The Prominence Tower
Jalan Jalur Sutera Barat No. 15, Alam Sutera
Tangerang, Banten 15143 - Indonesia

July 2019
Revision after the Thesis Defense on 17 July 2019

STATEMENT BY THE AUTHOR

I hereby declare that this submission is my own work and to the best of my knowledge, it contains no material previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any other degree or diploma at any educational institution, except where due acknowledgement is made in the thesis.

Timothy Dicky

Student Date

Approved by:

SWISS GERMAN UNIVERSITY

Alva Erwin, ST., M.Sc., MTI

Thesis Advisor Date

Ir. Heru Purnomo Ipung, M.Eng

Thesis Co-Advisor Date

Dr. Maulahikmah Galinium, S.Kom, M.Sc

Dean Date

ABSTRACT

DEVELOPING A SCALABLE AND ACCURATE JOB RECOMMENDATION SYSTEM WITH DISTRIBUTED CLUSTER SYSTEM USING MACHINE LEARNING ALGORITHM

By

Timothy Dicky
Alva Erwin, ST., M.Sc., MTI, Advisor
Ir. Heru Purnomo Ipung, M.Eng, Co-Advisor

SWISS GERMAN UNIVERSITY

The purpose of this research is to develop a job recommender system based on the Hadoop MapReduce framework to achieve scalability of the system when it processes big data. Also, a machine learning algorithm is implemented inside the job recommender to produce an accurate job recommendation. The project begins by collecting sample data to build an accurate job recommender system with a centralized program architecture. Then a job recommender with a distributed system program architecture is implemented using Hadoop MapReduce which then deployed to a Hadoop cluster. After the implementation, both systems are tested using a large number of applicants and job data, with the time required for the program to compute the data is recorded to be analyzed. Based on the experiments, we conclude that the recommender produces the most accurate result when the cosine similarity measure is used inside the algorithm. Also, the centralized job recommender system is able to process the data faster compared to the distributed cluster job recommender system. But as the size of the data grows, the centralized system eventually will lack the capacity to process the data, while the distributed cluster job recommender is able to scale according to the size of the data.

Keywords: Machine Learning, Distributed Cluster System, Hadoop MapReduce

© Copyright 2019
by Timothy Dicky
All rights reserved

SWISS GERMAN UNIVERSITY

DEDICATION

I dedicate this works for the future of the country I loved: Indonesia

SGU®
SWISS GERMAN UNIVERSITY

ACKNOWLEDGEMENTS

First and foremost, I grateful to God for His blessing and health He gave so that I am able to complete my thesis work.

My advisor, Mr. Alva Erwin, that inspires me to do this thesis work, for all his feedback and guidance he gave and I appreciate all of the expertise he shares.

My co-advisor, Mr. Heru Purnomo Ipung, for all encouragement he provides for me to finish my thesis work.

Not forget to mention my parents for their motivation support for me, encouragement, prayer and financial support they provide for me during this thesis work.

Last I want to appreciate all my friends and colleagues that keep me motivated to complete this thesis.

SGU ®
SWISS GERMAN UNIVERSITY

TABLE OF CONTENTS

Page

SWISS GERMAN UNIVERSITY

SWISS GERMAN UNIVERSITY

SWISS GERMAN UNIVERSITY