**CAREER ANALYSIS**

Submitted in partial fulfillment of the requirements for the award of degree of

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE & ENGINEERING**



**Submitted to:**

**DR. ROHIT BAJAJ**

**Submitted By:**

**ACHAL AWASTHI**

**(15BCS2142)**

**SHIKHA MAHAJAN**

**(15BCS2221)**

**ANKITA SINHA**

**(15BCS2240)**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**Chandigarh University, Gharuan**

**Jan-May 2018**

**ACKNOWLEDGEMENT**

I am highly grateful to the HOD CSE, Chandigarh University, Gharuan, for providing this opportunity to carry out the project. I would like to expresses my gratitude to other faculty members of Computer Science & Engineering department CU, Gharuan for providing academic inputs, guidance & encouragement throughout the training period. I would like to express a deep sense of gratitude and thank to Dr. Rohit Bajaj, without whose permission, wise counsel and able guidance, it would have not been possible to pursue my training in this manner. The help rendered by Dr. Rohit Bajaj, Supervisor for experimentation is greatly acknowledged. Finally, I express my indebtedness to all who have directly or indirectly contributed to the successful completion of my industrial training.

**ANKITA SINHA**

**ACHAL AWASTHI**

**SHIKHA MAHAJAN**

## CERTIFICATE

This is to certify that “**ACHAL AWASTHI**, **SHIKHA MAHAJAN**, and **ANKITA SINHA**” have developed the project entitled “**CAREER ANALYSIS”.** They have worked on the project **CAREER ANALYSIS.** They have used IBM Infoshere, VMWare, HADOOP famework, And PIG for the project. Their work is satisfactory.

I wish them all the best for their bright future.

**ABSTRACT**

Twitter, Facebook, Instagram, LinkedIn, and many more social media site receives millions of tweets, status, stories, and applications every day on variety of important topic. This large amount of raw data can be used for industrial, Social, Economic, Government policies or business purpose by organizing according to our need and processing. Hadoop is one of the best tool options for Social media data analysis and Hadoop works for distributed big data, Streaming data, Time Stamped data, text data etc. This project will analyze the Real Time Data using Hadoop and according which decision-making and strategy planning can be done easily.

**List of Figures**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Title** | **Page No.** |
| 1. | Waterfall Iterative Model | 6 |
| 2. | Data Flow Diagram Level 0 | 11 |
| 3. | Data Flow Diagram Level 1 | 12 |
| 4. | Data Flow Diagram Level 2 | 13 |
| 5. | Top 10 most commonly used tags in this data set (OUTPUT) | 14 |
| 6. | Average time to answer questions (OUTPUT) | 15 |
| 7. | Number of questions which got answered within 1 hour (OUTPUT) | 16 |
| 8. | Tags of questions which got answered within 1 hour (OUTPUT) | 17 |

**Table of Contents**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Topic** | **Page No.** |
|  |  |  |
| 1 | Introduction | 1 |
|  |  |  |
| 2 | Literature Reviews | 2 |
|  |  |  |
| 3 | Objective | 3 |
|  |  |  |
| 4 | Scope | 4 |
|  |  |  |
| 5 | Problem Statement | 5 |
|  |  |  |
| 6 | Software Development Methodology | 6 |
|  |  |  |
| 7 | System Requirement Specification | 7 |
| 8 | Data Flow Diagram | 10 |
|  |  |  |
| 9 | Output | 14 |
| 10 | Reference | 18 |
|  |  |  |