

**A
TRAINING REPORT
On**



BIG DATA USING HADOOP

Minor Project
For
Partial fulfillment of
Diploma Computer Science & Engineering
Session (2018-19)

Submitted By:
Bharat Sharma(163500018)

Submitted to:
**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
UNIVERSITY POLYTECHNIC
G.L.A UNIVERSITY, MATHURA**



DECLARATION

I hereby declare that the work, which is being presented and submitted to Department of Computer Science, University Polytechnic as a part of curriculum activity is an authentic record of my own work, carried within the premises of “GLA University”.

Date:-



ACKNOWLEDGEMENT

First and foremost, we would like to thank our Guide of this project **Mr. MRIDUL DIXIT** for her valuable guidance and advice. He guided us greatly to work in this project. His willingness to motivate us helped tremendously. Besides, we would like to thank the authority of GLA University for providing us with a good environment and facilities to complete this project. It gave us an opportunity to participate and learn about the **Data Science Project**. Finally, and honorable mention goes to our families and friends for their understanding and support in completing this project. Without help of those mentioned above, this project could not have been completed.

BHARAT SHARMA(163500018)

SIDDHARTH VERMA(163500060)



ABSTARCT OF PROJECT

This project is the based on Data Science and name Sentiment Analysis of tweets Data. This project is regards the Sentiment of people in the form of comment.

The peoples comments as they wish on the you tube and we are checks the polarity of the sentiment of different-different peoples that exist in the dataset. We are using dummy dataset we stored that data on the hadoop daemon name hadoop distributed file system(HDFS). Specially this project is deal with Big data and we know hadoop provide better solution of this type of problems. We used hadoop of for storing and processing the data.

Pig Latin languages used for refining and mining the data and after processing we stored final result on HDFS. This type of project is based on research so it beneficial for any enterprise or organisation. In this project we used various methodology for complete the project.

So project have a several phases, like data ingestion, data wrangling and data mining. These all are the phase or stages where this project is pass so that there are mandatory to deep knowledge of Data Science and their terminologies that common used during project complementation.

The Scope of this type of project is much more in every sector. Like in Business areas, politics areas and also any type of organisation and enterprises this type of project play wonderful role in these area to increases the growth of company, and capable to take decisions to organisation to their product failure and success.

This project also used in Natural Processing Languages to make languages to capable to learn the machine efficient so we can say that this type of project help to learning machine learning that behave as a human being. so there is lot of enjoyment me and my team form staring and end of the project.

TABLE OF CONTENT

S.NO	Page No.
1. Certificate	2
2. Acknowledgement	3
3. Declaration	
4. Abstract of Project	4
5. Chapter -1	5-7
1.1 Introduction	
1.2 Purpose	
1.4 Scope	
6. Chapter – 2	8-11
System Requirement Study	
2.1 Software Requirement	
2.2 Hardware Requirement	
7. Chapter – 3	12-18
Proposed Methodology	
3.1 Introduction	
3.2 Ingestion Process	
3.3 Wrangling Process	
3.4 Data Refining	
3.5 Mining Process	
3.6 Analysis Methodology	
8. Chapter – 4	19-23
4.1 Coding	

4.2 Screen Shots

9. Chapter – 5

24-25

5.1 Conclusion

5.2 Bibliography

LIST OF FIGURE

Fig. No.	Name of Figure	Page No.
1	Raw Data	12
2	Word Dictionary	13
3	Data stored on Hadoop	13
4	Positive rating of tweets	14
5	Negative rating of tweets	14