**SIX WEEK SUMMER TRAINING REPORT**

on

**DATA SCIENCE & MACHINE LEARNING**

Submitted by

**RANJITH SK, NIVASH G & DHARANESH V**

[**B.Tech (Computer Science Engineering) Hons] - LPU**

Under the Guidance of

**PRANAY SHARMA**

(Instructor for the course/project at IBM(All Soft Solutions))



**DECLARATION**

We hereby declare that we have completed our six weeks summer training at  **All Soft Solutions(IBM), Punjab** through online classesfrom **June 6** to **July 8** under the guidance of **Pranay Sharma.** We declare that we have worked with full dedication during this training and our learning outcomes fulfil the requirements of training for the award of degree of **B.Tech (Hons)**, Lovely Professional University, Phagwara.

**DHARANESH V**

**RANJITH SK**

**NIVASH G**

Date:15/07/2022

Expert from IBM(Teacher): **Pranay Sharma**

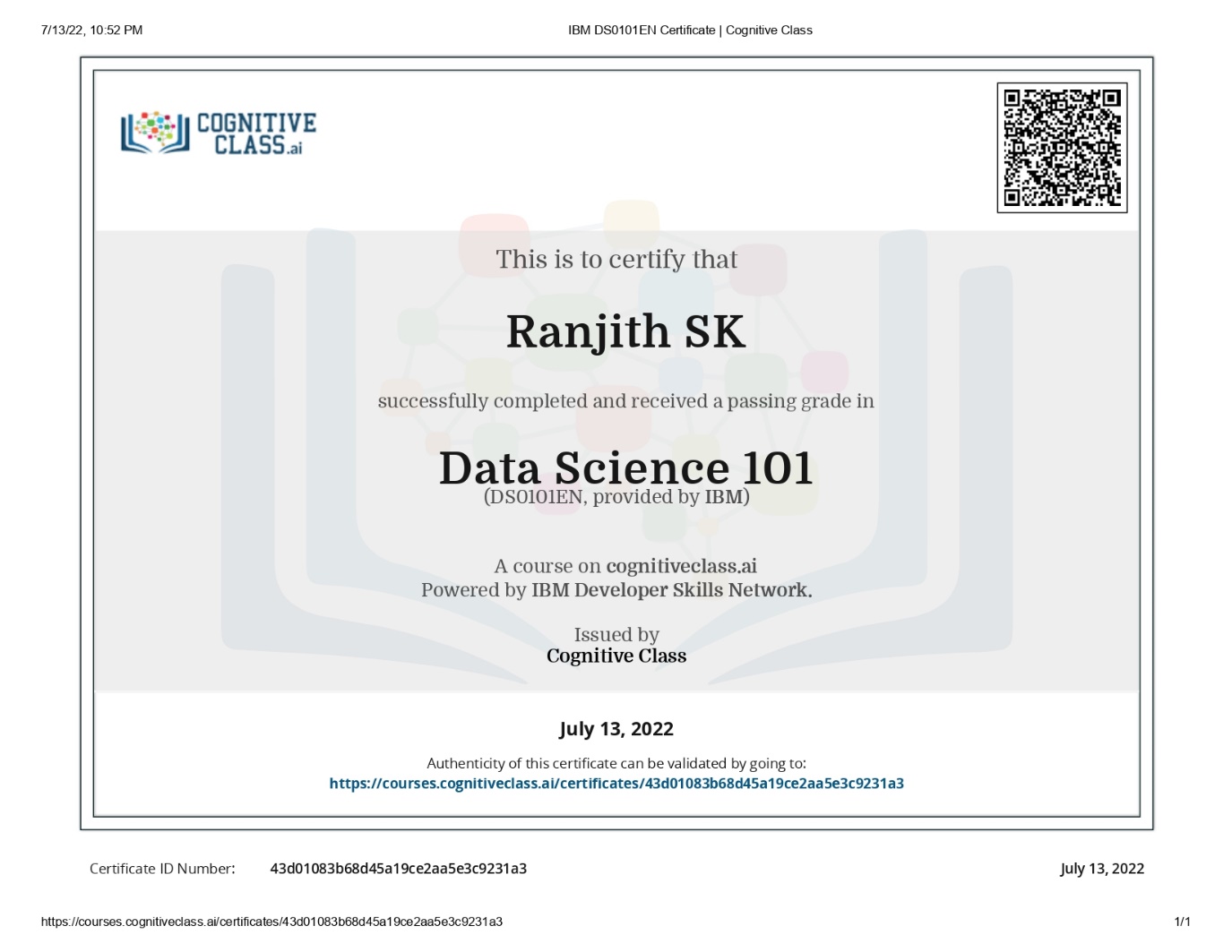
**ACKNOWLEDGEMENT**

The success and final outcome of this project required a lot of guidance and assistance from many people and we are extremely fortunate to have got this along the completion of our project work. Whatever we have done is only due to such guidance and assistance and we would not forget to thank them.

We would like to express our respect and gratitude to the teachers of Lovely Professional University and the coordinator for providing this opportunity to do this training program and project regarding **data science & machine learning.**

We would also like to extend our gratitude to **Pranay Sharma**(Teacher from IBM/All Soft Solutions) who helped us by giving guidance and support all through the program and also in completing the project.

**SUMMER TRAINING CERTIFICATE**



**TABLE OF CONTENTS**

1. Introduction
2. Keywords
3. Objective
4. Technology Learnt
5. System Implementation
6. Survey
7. Advantages of this Project
8. Reason for choosing this Technology
9. Implementation
10. Learning Outcome from training/technology learnt

**INTRODUCTION**

The project we have completed is based on the **analysis of WhatsApp chat and chat sentiment,** we have tried our best to make the complicated process of chat analysis as simple as possible using the necessary requirements. We have tried to design the project in such a way that user may not have any difficulty in using this package & further expansion is possible without much effort. Even though we cannot claim that this work to be entirely exhaustive, the main purpose of our work is Chat analysis which is the investigation and modelling of a specific WhatsApp chat’s thorough worktime and analytics, so that we can save our time and manage our worktime and activities.

The most used and efficient method of communication in recent times is an application called **WhatsApp**. WhatsApp chats consists of various kinds of conversations held among group of people. This chat consists of various topics. This information can provide lots of data for latest technologies such as machine learning. The most important thing for a machine learning model is to provide the right learning experience which is indirectly affected by the data that we provide to the model. This project aims to provide in depth analysis of this data which is provided by WhatsApp. Irrespective of whichever topic the conversation is based our developed code can be applied to obtain a better understanding of the data. The advantage of this project is that is implemented using simple python modules such as pandas, matplotlib, seaborn and sentiment analysis which are used to create data frames and plot different graphs, where then it is displayed in detail which is efficient and less resources consuming algorithm, therefor it can be easily applied to largest dataset.

This project is based on **data analysis and processing**. The first step in implementing a machine learning algorithm is to understand the right learning experience from which the model starts improving on. Data pre-processing plays a major role when it comes to machine learning. To make the model more efficient we need lots of data, so we turned our focus primarily on one of the largescale data producers owned by Facebook which is nothing but WhatsApp. WhatsApp claims that nearly 55 billion messages are sent each day. The average user spends 195 minutes per week on WhatsApp and is a member of plenty of groups. With this treasure house of data right under our very noses, it is but imperative that we embark on a mission to gain insights on the messages which our phones are forced to bear witness to.

**KEYWORDS**

WhatsApp chat data, Chat sentiment data, Pandas, NumPy, Seaborn, matplotlib, emoji, pillow, sentiment analyzer.

**OBJECTIVE**

In this decade the upcoming technologies are mainly dependent on **data**. This data can only be obtained if there is some research applied on the context of the requirements of the project/tool. Since a lot of machine learning enthusiasts develop models which helps solve multiple problems the requirements of appropriate data are very large scale this project aims to provide a better understanding towards various types of chats. This analysis proves to be better input to machine learning models which essentially explore the chat data. These models require proper learning instances which provides better accuracy for these models. Our project ensures to provide an in-depth exploratory data analysis on various types of WhatsApp chats.

**TECHNOLOGY LEARNT**

WhatsApp chat analysis is a **statistical analysis** project for WhatsApp chats. Working on the chat files that can be exported from WhatsApp it generates various plots showing, for example, which another participant a user responds to the most. We propose to employ dataset manipulation techniques to have a better understanding of WhatsApp chat present in our phones.

**SYSTEM IMPLEMENTATION**

**Python:** It is an interpreted, high-level general-purpose programming language. Created by Guido Van Rossum and first released in 1991. Its language constructs and objects-oriented approach aim to help programmer with clear, logical code for small and large-scale tools. Python is used for web development (server-side), software development, mathematics, it can be used alongside software to create workflows, it can connect to database systems, it can also read and modify files, it can be used to handle big data and perform complex mathematics and can be used for rapid prototyping, or for production-ready software development.

**Pandas:** This is an open-source Python libraries which is mainly used in Data Science and machine learning subjects. This library provides analysis tool for data manipulation, using its data structures this are used for analyzing data for manipulating time series analysis and numerical data.

**NumPy:** NumPy can be name come from Numeric Python, it is a data analysis library for Python that contains various numerical functions and methods for numerical analysis and also having multi-dimensional array objects and to process these arrays contains collection of routines.

**Matplotlib**: Matplotlib is easy to use and an amazing visualizing library in Python. It is built on NumPy arrays, and it work with the broader SciPy stack and consists of several plots like pie, line, bar, graph, scatter, histogram, etc. In this project, Matplotlib is used for various visualizations for analysis of whatsapp chats. Visualizations like bar charts, line charts, pie charts are used.

**Seaborn**: Seaborn is a library mostly used for statistical plotting in Python. To make statistical plots more attractive it provides beautiful color palettes and default styles. In this project, Seaborn is used for heatmap visualization for showing 24 hours with 7 day with different scale of color for getting hour with max to min messages.

**NLP:** In this project, Features of NLP are used like Parsing Text, Eliminating stop words and Analyzing Text. Parsing text is used for splitting messages into words for analysis like total words and mostly used words. A file is used that contains all stop words which is given to the python program to show meaningful words only by eliminating all stop words. Text analysis is used to identify how many media are shared; how many links are shared.

**SURVEY**

As a demo Survey analysis on the usage and Impact of WhatsApp Messenger, Various Studies and analysis has been done on the usage and impact of WhatsApp. Some of these studies are for finding the **impact of WhatsApp** on the students and some are based on for the public in a local region.

In a study of southern part of India was conducted on the age group of between 18 to 23 years to investigate the importance of WhatsApp among youth. Though this study, it was found that students spent 8 hours per day on using WhatsApp and remain online almost 16 hours a day. All the respondents agreed that they are using WhatsApp for communicating with their friends. They also exchange images, audio and video files with their friends using WhatsApp. It was also proved that the only application that the youth uses when they are spending time on their smart phone is WhatsApp. Methods used in this survey is to analyze the **intensity of WhatsApp usage** and its popular services and to identify the degree of positive or negative impacts of using WhatsApp.

**ADVANTAGES OF WHATSAPP CHAT ANALYSIS PROJECT**

• Works on all devices.

• Shows based on whatsapp chat file.

• Shows different visualizations.

• Total Messages.

• Total words.

• Media shared.

• Link shared.

• Monthly timeline.

• Most busy day.