GUJARAT TECHNOLOGICAL UNIVERSITY
(Established under Gujarat Act No. 20 of 2007)
ગુજરાતટેકનોલોજીકલ યુનિવર્સિટી
(ગુજરાત અધિનિયમ ક્રમાંકઃ ૨૦/૨૦૦૭ દ્વારા સ્થાપિત)

Annexure 1

Enrollment no: 190130111081

STUDENT'S WEEKLY RECORD OF INTERNSHIP

NAME OF STUDENT: Modi Shreshtha Pragnesh

DIARY OF THE WEEK: Dt: _____ **TO** _____

DEPARTMENT: Electronics And Communications Engineering **SEM:** 08

NAME OF THE ORGANISATION: Eternal Soft Solutions

NAME OF THE PLANT/SECTION/DEPARTMENT: Software and Cloud Engineering

NAME OF OFFICER INCHARGE OF THE PLANT/SECTION/DEPARTMENT: Mrs Poonam Patel

DESCRIPTION OF THE WORK DONE IN BRIEF

The following week I was introduced to the project that I will be working on. The project I will be working on revolves around securing your data even before it is released. My project involves anonymizing the user data using static and dynamic anonymization techniques so that the privacy of the user is not sacrificied and the data still remains statistically significant so that it can be used for research and analysis

My task is to use the raw, sensitive data and to build an end to end anonymization pipeline so that the user doesn't have to worry about inputting and anonymizing the data manually. This week was spent on brainstorming the solutions for building the pipeline and using the right tools to anonymize the data

My first hunch was to look at the solutions around AWS. I looked up services around compute and storage such as lambda, ec2, s3 and dynamodb. However, it so happened that I was told not to use s3 as s3 is used only for static data and the data in question would be updated fairly frequently and adding a lambda function on top of an ec2 instance for storing data would not be fruitful hence i had to look for other services

The next service that i came up with was AWS glue databrew which basically has all the data analysis and visualization functonality in the form of jobs and it had functonality to handle PII data. However, this was turned down because the data had quasi identifiers which could be linked with other data to get the original information and hence we needed a more sophasticated algorithm. Upon researching a little more, i found a static anonymizat ion tool called ARX which uses privacy models such as K-anonymity, L-diversity and t-closeness to anonymized your dara and hence it can not be traced back to the users. I downloaded the docker image for a local arx instance and got that running on my local machine and then I downloaded a module called pyarxaas which was a wrapper module providing python functions to work with the local arx machine and then i decided to use the data warehousing solution called snowflake to host my two seprate databases, conncet to snowflake using python and anonymize the data and generate the visualizations and this solution was approved

GUJARAT TECHNOLOGICAL UNIVERSITY
(Established under Gujarat Act No. 20 of 2007)

ગુજરાતટેકનોલોજીકલ યુનિવર્સિટી
(ગુજરાત અધિનિયમ ક્રમાંકઃ ૨૦/૨૦૦૭ દ્વારા સ્થાપિત)

TOTAL HOURS: 35

SIGNATURE OF STUDENT

✪ **The above entries are correct and the grading of work done by Trainee is**
EXCELLENT / VERY GOOD / GOOD / FAIR / BELOW AVERAGE / POOR

Signature of Faculty Mentor

Signature of officer-in-charge of Dept. / Section / Plant

Date:

Date:

✪ **Grading of Work, for trainee may be given depending upon your judgement about his Punctuality, Regularity, Sincerity, Interest taken, Work done etc.**

SUPPLEMENTRY NOTES
(add additional sheets if required)