



GUJARAT TECHNOLOGICAL UNIVERSITY

(Established under Gujarat Act No. 20 of 2007)

ગુજરાત ટેકનોલોજીકલ યુનિવર્સિટી

(ગુજરાત અધિનિયમ ક્રમાંક: 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

I spent majority of this week trying to understand how arx works and how to anonymize my data using ARX. Firstly, I installed the arx software and tried anonymizing a sample dataset following the instructions found in the arx documentation and then installed the local arx server using the github repository called 'Arx as A service' after installing the local docker image of the arx server, i ran it on the port 8080 and tried using the API to access the arx service.

However, the arx instance in question was a bit to messy to handle with the arx api and hence i decided to install the python module called pyarxaas. After several failed tries and debugging i realized that my system's current python version is not supported by the module as the module was outdated and hence i had no other option but to install the virtual manager module in python called pyenv as I could not downgrade my entire system's python version due to dependencies. After successfully installing pyenv, I installed python 3.8 and set it as the global version and installed pyarxaas in the python 3.8 and downgraded my numpy and pandas to match the version of pyarxaas

After installing pyarxaas, I used the very popular titanic dataset from kaggle and tried anonymizing it. Since the titanic dataset contains the names of the customer and the ticket number and the class of the passengers, it is an ideal candidate for anonymizing the data. I set some of the columns of the titanic dataset such as sex and embarked as insensitive. The column name has sensitive as the name data can easily be used to track back to the user. All the other columns such as Pclass, ticket and passenger id as quasidentifying as although they can not be used on their own to identify the data or the person but they can be used in conjunction with other columns or data to identify. I then built the hierarchies using pyarxaas and set the hierarchies using various privacy models such as kanonymity, ldiversity and tcloseness. I then identified the risk of reidentification and compared the risk of identification before the anonymization and after the anonymization



GUJARAT TECHNOLOGICAL UNIVERSITY

(Established under Gujarat Act No. 20 of 2007)

ગુજરાત ટેકનોલોજીકલ યુનિવર્સિટી

(ગુજરાત અધિનિયમ ક્રમાંક: 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.



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

ગુજરાત ટેકનોલોજીકલ યુનિવર્સિટી
(ગુજરાત અધિનિયમ ક્રમાંક: ૨૦/૨૦૦૭ દ્વારા સ્થાપિત)

SUPPLEMENTRY NOTES
(add additional sheets if required)

