BHASKARACHARYA NATIONAL INSTITUTE FOR SPACE APPLICATIONS AND GEO-INFORMATICS

WEEKLY PROGRESS REPORT (03/04/2023 – 09/04/2023)

WEEK 11

PROJECT NAME	MALWARE DETECTION USING ML

PROJECT DESCRIPTION:	**DESIGN AND IMPLEMENT ML MODEL TO DETECT MALWARE IN SYSTEM**
GROUP MEMBERS:	**SHASHANK SHARMA, PRATHAM PATEL, YASH SONI, SHUBHAM PATEL**
GROUP ID:	**I2**
GROUP GUIDE:	**HARSH KIRATSATA**
GROUP COORDINATOR:	**SIDHDHARTH PATEL**
COLLEGE GUIDE NAME:	**MAYANK PATEL, RITIKA LADHA, KINJAL CHAUDHARI**
COLLEGE GUIDE No.:	**+91 9033280445,8866438698, 9723173567**
COLLEGE GUIDE EMAIL.:	**mayank.patel@adaniuni.ac.in, ritika.ladha@adaniuni.ac.in, kinjal.chaudhari@adaniuni.ac.in**
COLLEGE NAME:	**ADANI INSTITUTE OF INFRASTRUCTURE AND ENGINEERING**

PROJECT LINK: https://github.com/shashankgsharma/malwaredetectionsystem

03/04/2023 TILL 09/04/2023 (7 DAYS)

Performing Traditional ML and MLP Deep Learning for behavioral analysis

03/04/2023	**Found Behavioral log dataset and performed Preprocessing and EDA on it**
04/04/2023	**Performed K-fold cross validation for train, test for 4 different traditional ml algorithms**
05/04/2023	**Comparison and EDA of results from traditional ml algos**
06/04/2023	**Researched about Neural Networks and sequential MLP model implementation**
07/04/2023	**Preprocessed data separately for MLP and trained the dataset for model development**
08/04/2023	**Did model building for MLP and evaluated the training and validation loss with increasing epochs**
09/04/2023	**Holiday**

WEEK 12(PLAN)	**We're planning to implement our model and test it on 4 other different unlabeled behavioral log datasets with different (malware):(benign) files ratio, and evaluate their results.**

REFERENCE:

- https://www.kaggle.com/code/davisl07/malware-detection
- https://github.com/rohan-paul/MachineLearning-DeepLearning-Code-for-my-YouTube-Channel/tree/master/Kaggle_Competition/Microsoft_Malware_Classification_BIG_2015
- https://github.com/dchad/malware-detection
- https://towardsdatascience.com/malware-detection-using-deep-learning-6c95dd235432
- https://viso.ai/deep-learning/deep-neural-network-three-popular-types/#:~:text=A%20multilayer%20perceptron%20%28MLP%29%20is%20a%20class%20of,computing%20power%20required%20by%20modern%20deep%20learning%20architectures

SCREENSHOTS:

1. TRADITIONAL ML:

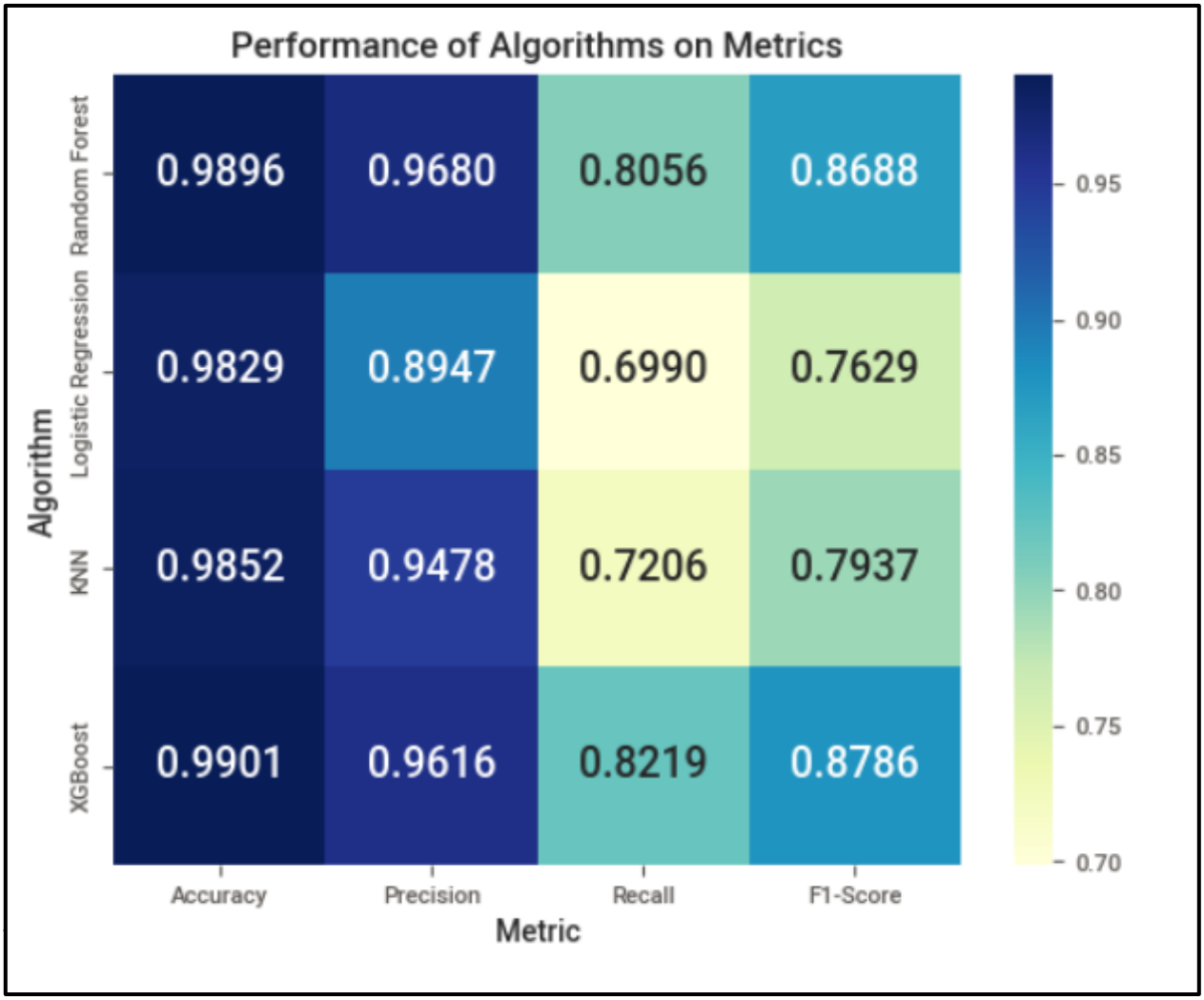
```
In [3]: data = pd.read_csv('C:/Users/Dell/Desktop/Sem8/dataset/dynamic_api_call_sequence_per_malware_100_0_306.csv')

In [4]: data.head(15).style.background_gradient(cmap='turbo')
```

Out[4]:

	hash	t_0	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9	t_10	t_11	t_12	t_13	t_14	t_15	t_16	t_17	t_18	t_19	t_20	t_2
0	071e8c3f8922e186e57548cd4c703a5d	112	274	158	215	274	158	215	298	76	208	76	172	117	172	117	172	76	117	35	60	81	6
1	33f8e6d08a6aae939f25a8e0d63dd523	82	208	187	208	172	117	172	117	172	117	172	117	172	117	172	117	172	117	172	117	172	11
2	b68abd064e975e1c6d5f25e748663076	16	110	240	117	240	117	240	117	240	117	240	117	240	117	172	117	99	260	141	65	240	11
3	72049be7bd30ea61297ea624ae198067	82	208	187	208	172	117	172	117	172	117	172	117	172	117	172	117	172	117	172	117	172	11
4	c9b3700a77facf29172f32df6bc77f48	82	240	117	240	117	240	117	240	117	172	117	172	117	16	240	117	11	274	158	215	274	15
5	cc6217be863e606e49da90fee2252f52	117	208	117	208	117	240	117	240	117	208	228	215	274	158	215	274	158	215	240	117	71	29
6	f7a1a3c38809d807b3f5f4cc00b1e9b7	215	274	158	215	274	158	215	172	117	172	117	172	117	198	208	260	257	25	240	117	99	2
7	164b56522eb24164184460f8523ed7e2	82	240	117	240	117	240	117	240	117	240	117	172	117	172	117	16	31	86	112	271	111	8
8	56ae1459ba61a14eb119982d6ec793d7	82	240	117	240	117	240	117	240	117	240	117	16	208	187	208	240	117	39	35	171	172	11
9	c4148ca91c5246a8707a1ac1fd1e2e36	82	208	187	208	172	117	172	208	16	208	240	117	240	117	82	112	123	65	112	123	65	26
10	fb7569d1c2c1fa36a97fdc732f51a637	172	117	208	76	274	158	215	274	158	215	76	215	76	172	117	172	117	286	240	286	297	13
11	e7ac6a2de45506164777941faf953094	82	240	117	240	117	93	117	172	117	16	117	215	228	208	240	117	82	198	86	82	274	3
12	1282837376a698e38af5cca54bdfbdd0	82	172	117	16	294	94	215	274	158	215	274	158	215	94	208	274	158	215	274	158	215	27
13	2688d03495ba17054a9a65028a0a80f8	82	240	117	240	117	240	117	240	117	172	117	172	117	16	240	117	11	274	158	215	274	15
14	2109cd66383a81926aef367530a2a9fc	82	240	117	240	117	240	117	240	117	172	117	172	117	16	240	117	11	274	158	215	274	15

```
In [5]: data.shape
```

Out[5]: (43876, 102)



2. DEEP LEARNINIG MLP MODEL: