Field Worker's Routine Behaviour for Efficient Time Utilization

Munish Mehta, Rahul Sharma, Pawan Suthar

Abstract: ***In this paper, we are proposing a monitoring technique that can predict the appropriate time to meet the field officer with more accuracy as compared to general human prediction. To get the routine behaviour of an individual we will keep a track of his online presence on WhatsApp by the "Active Now" status shown beneath the user's name in the profile. This will help the Personal assistants, secretaries and receptionists to provide right appointment schedule to clients or to those who want to meet a field officer. This all will be done by scraping data from WhatsApp web portal using python modules, storing the judging parameters like unique identification number, date and time stamp, duration of continuous presence and number of sessions in a day, afterwards analyzing the dataset and providing an appropriate routine behaviour prediction. For the sake of efficient utilization of time and resources.***

Index Terms**: Human behaviour analysis, Naïve Bayes, Time Utilization, Web scraping.**

I. INTRODUCTION

Today, in this modern world that is being governed by the digital devices which are making our communications much easier and faster than ever. Our tendency to check the online presence and be connected on the handheld devices has increased tremendously. The advancement in communication technology to the next level and has allowed humans to express their thought and well about thoughts and other information easily. Almost every large-scale organization has its own dedicated application software through which they control their entire production, manufacturing, billing and other internal processes. This sophisticated designed software popularly known as portals, help the supervisors to take the important decisions very efficiently and also to share the data and all the information among all the employees very effectively. Now almost every software has a sign in or login section through which any employee in the organization gets access to this system and start working on that. Time is the most valuable asset in an individual's life and how a person spends it throughout the day can be a judging factor of one's personality traits. So the amount of time a person spends on social media like WhatsApp shows how much effect it is doing to one's productivity during the work hours. In various government departments, to meet a field officer has been a tedious task as they might not be present in their office from time to time. If anybody asks for an appointment to meet them it is unpredictable to say an appropriate time period.

In today's technological world, almost every organization has a need to monitor the resources and the employees so that they can find the optimum way to take up the decisions and we know that there are dedicated software available in the market for different type of industries but there is not any generic software available in public domain which can be deployed in any organization and also which support the monitoring even with popular social media platforms. Our system can monitor the activity of any employee who is accessing the social media platform like WhatsApp, Facebook etc. and hence it is very useful for any organization. For example, last month we were thinking of creating an index of the sentiment of students in various classes about what they actually feel about the class, were they interested in the content that was being taught in the class? Was the teacher able to express what he/she trying to convey to them?

II. LITERATURE SURVEY

In [4], authors have taken the data of the sensors from a mobile system and then analyzed that data. But as we can see in our neighbourhood that many people use more than one device throughout their day time as most of the working person use their office systems and sometimes uses their smartphone for messaging purposes and all. So Just by getting the data from one device is not sufficient for accurately analyzing the daily routine and human behaviour. Reference [9] tells us about the work that has been done with various aspects of mobile crowd sensing to predict human behaviour and their routine habits.

Commonly, people in the research field of this domain try to assert the behaviour of a person by a particular feature's data but that might not be sufficient for best judgment on one's behaviour. Reference [5] shows that the portals have now support for mobile devices and some have android applications so employees can be connected through their mobiles as well.

Previous studies that are done using the sensors of the particular mobile for behaviour analysis can be improved because a person can be using a computer at one point of time and mobile on another point of time.so our system support multiple devices and platforms and also the consumer participation and extra sensoring would be minimal. Reference [6] describes how we can classify the routine of

Revised Manuscript Received on July 02, 2019.

Munish Mehta, munish.mehta80@gmail.com, Department of Computer Application, National Institute of Technology, Kurukshetra, India.

Rahul Sharma, rahul_51710031@nitkkr.ac.in, Department of Computer Application, National Institute of Technology, Kurukshetra, India.

Pawan Suthar, pawan.suthar75@rajasthan.in, Jodhpur Discom, Electricity Department, Government of Rajasthan, Bikaner, India.

a person according to the human behaviour models that we are trying to develop right now. These paper will help us to build the behavioural models.

The system would be used for predicting the time when a person leaves the portal and sign-in in the portal and for how much time does he continue to do work on the portal and etc.

This tool can also be deployed on the social media platforms considering them as a dedicated web portal.

This can also be useful in predicting the behaviour or wakeup and sleep time of a particular person, helpful to know how prone one is to health issues. To fetch and extract the data from the portals to analyze the routine behaviour of a field officer we will use the Crawling Methodology described in [7] with the application framework provided. To crawl an entire website for only particular data is not a good practise so we will use the optimised technique given by Jing Wang et. al [7] to extract the required data using Scrapy framework and then converting it into data in CSV file format using pandas Framework.

III. INPUT SOURCES & TECHNOLOGY USED

Our tool can be used on data and services in many types of organizations whether it's a PSU, govt. Organization or a private sector firm. This will help the organizations to make optimum utilization of human resources by monitoring their activities and work. It will also help them to deduce timings for breaks in between the work hours, or to make an appointment for persons who want to meet any of the employees. Reference [8] gives us an imaginary view on how to make a user's daily routine less time wasting by collecting one's activity stats using multiple wearable sensors that can track the routine behaviour and predict one's habits. This system of Wearable sensors can tell whether the person is standing, running, walking, eating at home or eating on the street corner, talking or working on a desk in the office.

This tool can be used in an educational institute to look over its students and teachers, or in a firm that has employees working as customer support executive to keep a check over them, and many more scopes can be found as the tool enhances as it gets spread to the open-source community.

A. Python modules and libraries used

We will be using some distinct Python library modules[1] to make the smart use of functions that are once defined for future use, some of them are named and explained below:

Reference [3] shows that Beautifulsoup is a Python module that is used to fetch information by pulling data out of XML and HTML files. This module works with various basic standard python library parsers but it can also make use of third-party parsers of navigating, searching and modifying the parse tree. It commonly saves programmers hours or days of work.

The *urllib2* python module defines functions and classes that are used to perform some basic actions on URLs which help in linking with the URLs (mostly HTTP) in a complex world -- basic and digest authentication, etc.

The *Panda* python module provides efficiently fast data frame objects with integrated indexing to manipulate data. It aims to be the fundamental high-level building block for doing practical, real-world data analysis in Python. It also provides a tool for reading and writing data from-to pre-defined data structures and other file formats like- CSV, text, Excel sheets, SQL etc.

NumPY is the fundamental python library module for the numeric functions of mathematics this allows us to make use of scientific computing and is explained in [2].

Matplotlib is a python library that is used to convert the numeric input to the graphical representation in two dimensions and it then plots the output on the charts using python scripts. This Python module converts the analyzed data into a chart representation structure.

Selenium is a tool that when integrated with a language can help to automate the web browser and test the web applications. It requires a driver to create an interface with the selected browser. For Firefox it requires *geckodriver* and for the chrome we need *chromedriver*.
chromedriver uses the python script to automate the web browser that initiates the scraping of the data. Then to scrape the actual content we used *Scrapy* application framework which is designed in python language along with The Twisted, which is a networking framework and is event-driven.

B. Web Server

For *Apache Tomcat 6.0* is a platform that provides a servlet container and HTTP server environment to the web-based applications. It is developed by Apache Software Foundation and is an open-source Java servlet container.

C. Algorithm To Be Used

To analyze the collected data of a target individual from a social media web portal, WhatsApp for instance here, we are going to use the Naïve Bayes algorithm for our project. In probability theory, With respect to two random events, Bayes' theorem relates the conditional and marginal probabilities. It is often used to compute posterior probabilities given observations. A naive Bayes classifier is a term dealing with a simple probabilistic classification based on applying Bayes' theorem. In general terms, a Naïve Bayes classifier assumes that the presence (or absence) of a specific feature of a class is not related to the presence (or absence) of any other feature.

D. Data Collection And Preprocessing

In this paper, a famous communication application known as WhatsApp is chosen as our sample application portal for the whole project as it is widely used by many people and also for accurately analyzing one's daily routine, this application suits the best till now. So first of all, we would have to collect the data from the WhatsApp for further steps. And as we got to know that WhatsApp doesn't have any official Application Programming Interfaces(APIs) for data scraping from its application. But the WhatsApp runs on both mobile application platform and on the computer as a user can use it through their website

web.whatsapp.com. So we used selenium, [10]Beautifulsoup and Scrapy framework to automate and scrap our required data from the web portal. So now our next step is to find what elements from the web portal are appropriate for us to be scrapped. After the selection of appropriate features for analysis, we used the CSS filter function to filter the scraped data using extract text method of the scrapy framework.

Data items that were scrapped from the website of WhatsApp with the consent of that particular person.

1. The time stamp at which the user was online.
2. His mobile no. for its unique identification.
3. WhatsApp cookies if any.

Here we used beautifulsoup to parse the HTML code and then find each of the elements from the HTML tree. Afterwards, it saves the value of each item in a CSV file for further steps.

E. Processing and organizing of data

Every day has different CSV file consisting of all the timestamps and time intervals when that particular person was online, So the first question here is how to divide the data received either month wise or week wise. So we are dividing and then analysing the data week wise.

Table 1 – CSV file representation

days	su	1	0	1	1	1	1	0	0	0	0
	mo	0	1	1	1	1	0	0	0	1	1
	tu	0	1	1	1	1	1	0	0	0	0
	we	1	1	1	1	1	1	1	1	0	0
	th	1	1	1	0	0	0	1	1	1	0
	fr	0	0	1	1	1	0	0	1	0	0
	sa	0	0	1	1	1	1	1	0	0	0
hours		8	9	10	11	12	13	14	15	16	17
Timestamps at certain intervals (every hour)											

In Table1, 0 depicts Offline and 1 depicts Online availability of the user. All the data here is assumed and has no relevance right now with any actual survey. The details regarding the person will be observed on an hourly basis for all days of a week. The key attributes that we will consider are shown with the help of the table in Table 2 :

Table 2 – Collected data for active sessions of the user

Week Day	Wake Up Time	Start time of Longest Session	End time of longest session	Max Duration of session in hrs	Last checking time
Monday	09:00	09:00	12:00	3	17:00
Tuesday	09:00	09:00	13:00	4	13:00
Wednesday	08:00	08:00	15:00	7	15:00
Thursday	08:00	08:00	10:00	2	16:00
Friday	10:00	10:00	12:00	2	15:00
Saturday	10:00	10:00	14:00	4	14:00
Sunday	08:00	10:00	13:00	3	13:00

The collected wake-up time, start and end time of maximum duration of the active session were in separate tables for analysis. The Graph for the collected data can be depicted as the following graph in Fig. 1

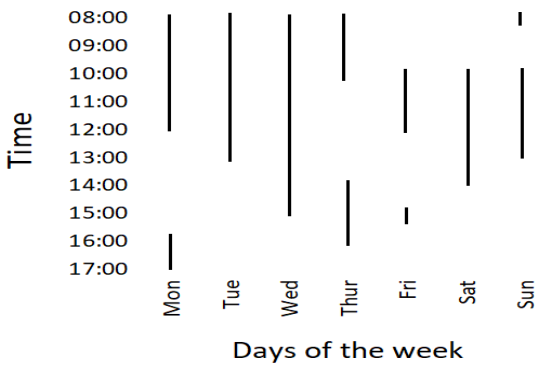

Fig.1 – Graphical Representation of Collected Data

We are mainly finding the following characteristics by analysing the graph of a particular person.

1. His mean wakeup time or the time when he first checked his WhatsApp app either from mobile or pc for the whole week.
2. Calculating Longest time interval in the whole day when he was online on that particular portal continuously.
3. His ending time or last time period when he checked his account.

For calculating the predicted wakeup time for a person for a week, we would use the following formula:

$$Median = \{(n+1) \div 2\}th\ term \quad (1)$$

The collected data of user's activity throughout the week is to be plotted on the graph for a better understanding of time prediction for everyday awakening time and his/her last login time. It's much easy to check the variation of daily activities' time through a graph.

Fig. – 2 and Fig. – 3 shows a clear picture of variation.

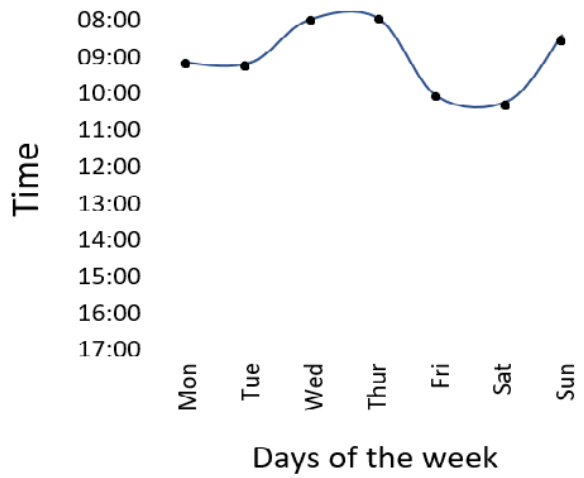

Fig. – 2 Graph of Initial Check-in for the day

Fig. – 2 shows that the user checked the WhatsApp initially at 9 am, 9 am, 8 am, 8 am, 10 am, 10 am, 8 am on Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday respectively.

Similarly, Fig. – 3 describes the last checking time of user on the WhatsApp portal. Providing a Pictorial form of variation in activity tracks.

F. Medium Of value in Dataset(CSV file)

Here we used beautifulsoup to parse the HTML code and then find each of the elements from the HTML tree. Afterwards, it saves the value of each item in a CSV file for further steps. Graphs in Fig.- 1, 2, 3 would be plotted from the data of the CSV file which will be scraped using python library modules. The Fields accordingly will help to plot the graph for each attribute.

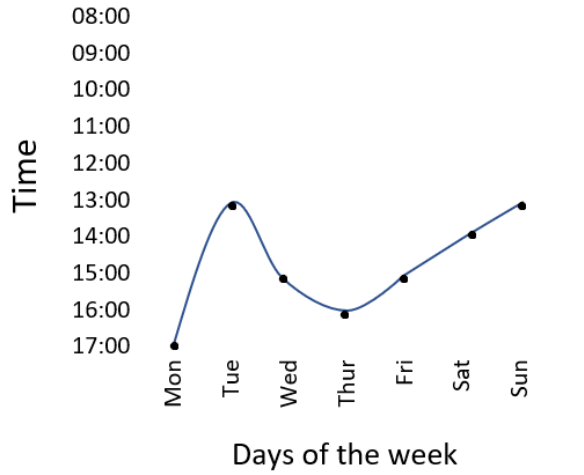

Fig. – 3 Graph of Last Check-in for the day

IV. CONCLUSION

As we all know due to the frequent use of technology in every domain the need of monitoring the data activity as well as the activist has become almost mandatory. For an institution to continue operating successfully and productively monitoring at every level like network, equipment and systems even the employees is a key work. Problems, faults and failure of modules may rise into critical issues. This tool will help them to reduce the wastage of time of a client. This analysis technique can then be used in various other domain like Equity share market to judge daily hourly basis trend of its price, attentively listening consistency of a student in class during the lectures and many more.

Therefore, in every work field where digital infrastructure is primary, it is the first priority to keep an eye on it. And make it sure that the services rendered to the end user are up to the mark. Secondly, it can be used by an educational institution to monitor their staffs' as well as students' activity during the working hours, for example, it can monitor the students and the lecturer without letting them know on social networks or on their own network if it exists, other than that it can predict the appropriate time to meet someone by collecting data on their routine they do on daily, weekly and monthly basis. This can be done by analyzing the data and converting it into graphs and charts that can predict the most appropriate time to meet one without wasting one's precious time.

REFERENCES

[1] Python Reference. [Online]. Available: www.python.org
[2] (Basic Book) Al Sweigart. (2015,April,14). Automate the boring stuff with PYTHON (Second Edition). Available: https://automatetheboringstuff.com/
[3] Leonard Richardson. (2004-2015). Crummy, The Webspace by L.R. [Online]. Available: https://www.crummy.com/software/BeautifulSoup/bs4/doc/
[4] Mafrur, R., Nugraha, I.G.D. & Choi, D. Hum. Cent. Comput. Inf. Sci. (2015) 5: 31. https://doi.org/10.1186/s13673-015-0049-7.
[5] Harari G.M., Muller S.R., Aung M.S., Rentfrow P.J. Smartphone sensing methods for studying behaviour in everyday life (2017) *Current Opinion in Behavioral Sciences*, 18, pp. 83-90.
[6] Nikola Banovic, Tofi Buzali, Fanny Chevalier, Jennifer Mankoff, Anind K. Dey, Modeling and Understanding Human Routine Behavior, Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, May 07-12, 2016, Santa Clara, California, USA [doi>10.1145/2858036.2858557].
[7] Jing Wang, Yuchun Guo Sch, "Scrapy-Based Crawling and User-Behavior Characteristics Analysis on Taobao", 2014 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, pp. 44-52.
[8] Mizuno, H & Nagai, H & Sasaki, K & Hosaka, H & Sugimoto, Chika & Khalil, K & Tatsuta, S. (2007). Wearable Sensor System for Human Behavior Recognition (First Report: Basic Architecture and Behavior Prediction Method). 435 - 438. 10.1109/SENSOR.2007.4300161.
[9] Jian, An & Xiaolin, Gui & Jianwei, Yang & Yu, Sun & Xin, He. (2015). Mobile Crowd Sensing for Internet of Things: A Credible Crowdsourcing Model in Mobile-Sense Service. Proceedings - 2015 IEEE International Conference on Multimedia Big Data, BigMM 2015. 92-99. 10.1109/BigMM.2015.62.
[10] Ashiwal Pratiksha, Tandan S.R., Tripathi Priyanka, Miri Rohit, Web Information Retrieval Using Python and BeautifulSoup, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 4 Issue VI, June 2016, Available: https://www.ijraset.com/fileserve.php?FID=4999

AUTHORS PROFILE

Dr Munish Mehta, Assistant Professor at NIT Kurukshetra Doctor of Philosophy (PhD), Master of Computer Application.

Rahul Sharma MCA Student at NIT, Kurukshetra, Bachelor of Computer Application from the University of Rajasthan.

Pawan Suthar working as an Informatics Assistant in Jodhpur Discom, Government of Rajasthan, Bachelor of Computer Application from Maharaja Ganga Singh University, Bikaner.