A Machine Learning Based Mobile Data Recommendation system

[1]Jeevan kumar, [2] Pankaj Pandey
[1] Research Scholar, Oriental Institute of Science and Technology, Bhopal
[2] Associate Professor, Oriental Institute of Science and Technology, Bhopal

ABSTRACT

*Quality of Service (QoS) is one of the most important success criteria for Internet Service Providers (ISP) who own network infrastructures, such as transmission media for both wired and wireless networks, speed of communication(bandwidth) to the Internet, etc. But As a user, we need to advance resource management to maximize the utilization of our resources according to our requirements. Internet Service Providers (ISP) provides only a fixed amount of data according to plan purchase by the user with maximums possible speed(bandwidth) of the user network circle. To use the data pack by device, need an advanced system to manage all data smartly. The main goal of this work is to combine both analysis user needs and analysis user data to find the most possible accurate path to use data. The device calculates the most accurate way to use all data and automatically manage uses speed (bandwidth) to after the end of the package duration data full uses. Today we perform most of the work help of internet in a recent survey we see the use of the internet in per days increase rapidly and Internet Service Providers (ISP) in order to improve their products and services and provide day to day more speed(bandwidth) and reliable service so we need very carefully to use our data pack otherwise we consume data in very less time some time it's not beneficial .Now a day's researchers dealing with how to improve the infrastructure of network and how to provide more bandwidth to our user so they feel the better experience. Present days Internet Service Providers (ISP) to control Bandwidth use a static algorithm. In this time Internet Service Providers (ISP) some statics technique to control bandwidth like after use of 80 percent of they reduce speed to fixed basics speed some ISP not reduce the speed they continue till data pack, not complete. Now this day for a fixed plan of data pack we need a smart Recommender Systems to manage speed (Bandwidth) to proper use of all available data according to everyone need.***

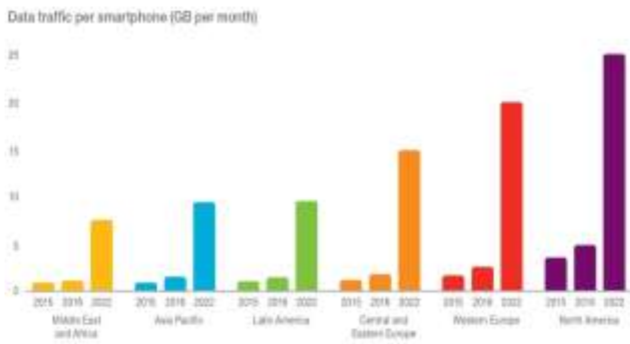
Keywords-*: Internet Service Providers, Recommender Systems, Bandwidth, Data Mining,A priori, Classification*,.

1.1 Introduction

Computer information will be data prepared or put away by a PC. This data might be as content reports, pictures, sound clasps, programming projects, or different kinds of information. PC information might be prepared by the PC's CPU and is put away in records and envelopes on the PC's hard plate. At its most simple level, PC information is a lot of ones and zeros, known as paired information. Since all PC information is in a parallel arrangement, it tends to be made, handled, spared and put away carefully. This enables information to be moved to start with one PC then onto the next utilizing a system association or different media gadgets. It likewise doesn't break down after some time or lose quality in the wake of being utilized on different occasions [1].

Total mobile data consumption rapidly increasing all over the world and is predicted to reach 131 exabytes (EB) per month at the end of 2023. It means 2018 to 2024 compound annual growth rate (CAGR) 30% increment. Swedish telecom leading company Ericsson said in its mobility report. The monthly data need on every smartphone user increases five times from 2017 to 2023. According to this report, every Indian average data need in 2017 is 3.9 times they increase to 2023 to 18GB.

According to the Ericsson company report, 35 percent of the mobile data traffic will be carried by 5G networks. Smartphones continue to generate most of the mobile data traffic close to 90 percent today and projected to reach 95 percent at the end of 2024. In this forecast period, this means that only new or evolved smartphone-based services are likely to have the ability to significantly affect the global traffic growth curve.



Mobile data traffic by region	2016(GB/Month)	2022(GB/Month)
Asia Pacific	3.6	8
CEMA	1.2	12
Western Europe	1.2	8
Latin America	0.7	8
North America	1.8	6

Figure1.1 Data Traffic per smartphone

a) **Key highlights**

- Between 2016 and 2022, the traffic generated by smartphones will increase by 10 times
- By 2022, there will be 12 times more mobile data traffic in Central & Eastern Europe and Middle East & Africa (CEMA)

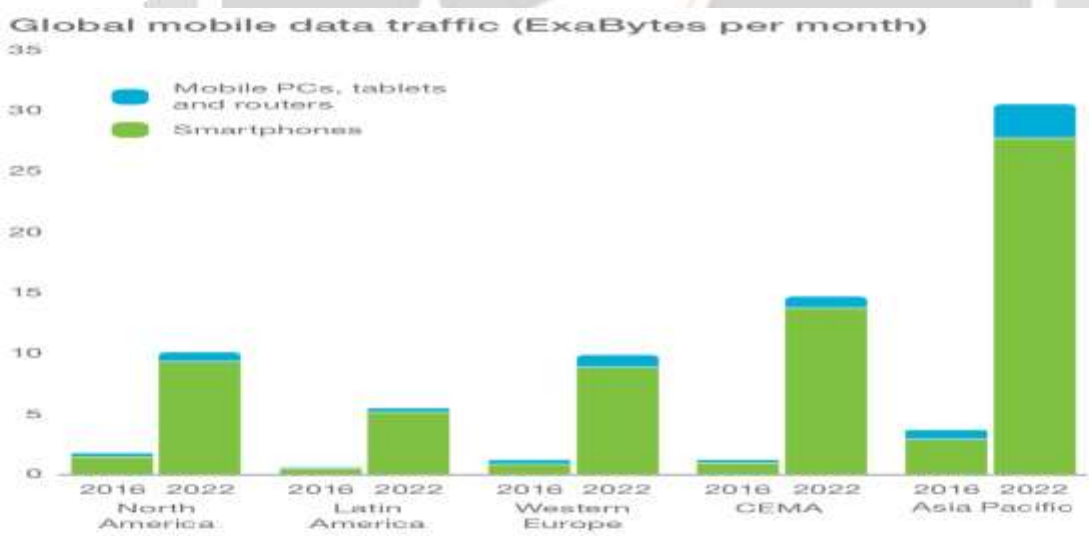


Figure1.2 Global Mobile Data Traffic

b) **Key highlights**

- Growth of smartphone data traffic compare computer tablets and routers rapidly grow in all region.
- All region computer tablets and routers data consumption almost double while smartphones data consumption growth more than five times increment [4].

1.2 Indian region data consumption report-

- A global innovative leader company Nokia recently published a report about future data consumption in India.
- Mobile data consumption 144 percent increase 2360 petabyte with every 4G user average data consumption 11Gb in 2017 to 2018.
- Overall, India data traffic grew by 144 per cent. 4G usage reaching 11 GB per user per month on average. Video content contributed up to 65 per cent of total mobile data traffic.
- The average consumption over both Wi-Fi and mobile networks in India was 8.8 GB data per user per month, at par with other developed markets, as per the report.
- Mobile broadband performance in India shows that 4G emerged as the key driver of mobile data consumption in 2017, capturing 82 per cent of mobile data traffic and growing 135 per cent (y-o-y) with the rapid deployment of 4G networks [2].

c) 1.5 India Telecom statistic bulletin report 2018

year wise data use per subscriber report-

Year	GSM(MB)	CDMA(MB)
2014	53.94	176.24
2015	89.06	278.22
2016	133.87	433.64
2017	1006.00	473.00
2018	2447.00	173.00

Above report clear show rapidly grow of data need between 2016 to 2018. In grow of this data need one important parameter evolution of network 3G to 4G. The 4G services start internally 27 December 2015 but officially company commercial plan launched in 5 September 2016.

1.3 Evolution of mobile network: The cell remote age (G) for the most part alludes to an adjustment in the essential idea of the administration, non-in reverse perfect transmission innovation, and new recurrence groups. New ages have shown up in at regular intervals since the principal motion picture from 1981-A simple (1G) to simple (2G) organize. After that there was (3G) sight and sound help, spread range transmission and 2011 all – IP Switched systems (4G) comes. The most recent couple of years have seen a wonderful development in the remote business, both as far as versatile innovation and its supporters. There has been an unmistakable move from fixed to versatile cell communication, particularly since the turn of the century. Before the finish of 2010, there were more than multiple times more portable cell memberships than fixed phone lines. Both the versatile system administrators and merchants have felt the significance of proficient systems with a similarly effective plan. This brought about Network Planning and improvement related administrations coming into sharp core interest. Cutting edge portable systems usually alluded to as 4G, and are visualized as a large number of heterogeneous frameworks interfacing through a level IP-driven engineering. The 5G center is to be a Re-configurable, Multi-Technology Core. The canter could be an assembly of new innovations, for example, Nanotechnology, Cloud Computing, and Cognitive Radio, and dependent on All IP Platform. These new innovations necessities represent a few difficulties toward 5G advancement. Portable Cellular Network development has been sorted into 'ages' as:

1.4 Recommender Systems Overview

- Recommender systems are techniques and tools providing a recommendation to users.
- The Recommender system provides useful and practical advice, usually customized according to user preferences or preferences.

- The purpose of the Recommender system (RS) is to provide users with meaningful suggestions about the data consumption or control in data use they might reduce data consumption.
- This new area of research is gaining more importance mainly due to the effects of widespread use of mobile-based service.
- The purpose of this system to provide a suggestion-based estimate of the use of user data and total need for data of every specific user according to his per day use log basis.

1.4.1 Classification of Recommender System

The recommender system is to provide users with the recommended techniques and tools. We often talk about recommending articles to users, where the articles are generic terms that are used to represent what the system recommends to users. The purpose of the recommender system is to provide useful and practical advice, generally adapted to the user's preferences or tastes. [8].

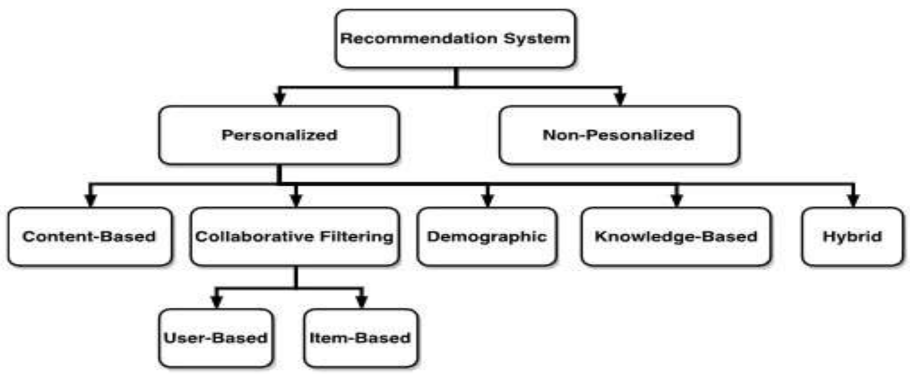


Figure 1.2: Recommender System Classification

In general, we can classify recommendation algorithms in different types based on the technology that produces the recommendations. According to [9], we will classify the recommender system based on the technology that produced the recommendation. For clarity, Figure 1.1 shows the classification.

1.5 Basket Analysis Algorithm

The total calculation can be isolated into two stages:

Stage 1: Apply the least help to discover all the successive sets with k things in a database.

Stage 2: Use oneself join principle to locate the incessant sets with k+1 thing with the assistance of successive k-itemset. Rehash this procedure from k=1 to the moment that we can't make a difference oneself join rule.

This methodology of broadening a regular itemset each one, in turn, is known as the "base up" approach.

Mining Association Rules

Till now, we have taken a gander at the A priori calculation as for regular itemset age. There is another assignment for which we can utilize this calculation, i.e., discovering affiliation governs effectively. For discovering affiliation rules, we have to discover all standards having to bolster more noteworthy than the limit backing and certainty more prominent than the edge certainty. Be that as it may, how would we discover these? One conceivable way is beast power, i.e., to list all the conceivable affiliation leads and compute the help and certainty for each standard. At that point dispose of the standards that bomb the edge backing and certainty. However, it is computationally substantial and restrictive as the quantity of all the conceivable affiliation standards increments exponentially with the number of things.

Given there are n things in the set I, the all-out the number of conceivable affiliation principles is 3n–2n+1+1.

We can likewise utilize another way, which is known as the two-advance methodology, to locate the productive affiliation rules.

The two-advance methodology is:

Stage 1: Frequent itemset age: Find all itemset for which the help is more prominent than the edge bolster following the procedure we have just observed before in this article.

Stage 2: Rule age: Create rules from each continuous itemset utilizing the paired parcel of incessant itemset and search for the ones with high certainty. These guidelines are called competitor rules.

Give us a chance to take a gander at our past guide to get a proficient affiliation rule. We found that OPB was the successive itemset. So, for this issue, stage 1 is as of now done. In this way, let' see stage 2. All the potential standards utilizing OPB are:

OP⟶B, OB⟶P, PB⟶O, B⟶ OP, P⟶OB, O⟶PB

On the off chance that X is a successive itemset with k components, at that point there are 2k−2 applicant affiliation rules.

We won't go further into the hypothesis of the A priori calculation for principle age [19, 20].

Association score-Association score describes the score of all the associated parameters; they show all the association possible with a score. The help of python Pandas library method we arrange in ascending and descending order and then we choose the best score association rule according to this association we easily estimate data and apply data cap.

2. Problem Statement

The main goal of this work is analysis user streaming data needs and analysis user streaming data to find the most possible accurate way to use streaming activity. Our proposed model calculates the most accurate way to use all data and automatically manage speed (bandwidth) until data pack end during this bandwidth same and data pack also full uses. To provide a streaming recommender system based on Basket analysis (A priori algorithm) to find accurate way use of data on the basis of user previous streamed data consumption and user need streaming data. The proposed recommender system estimates the user data consumption on per day basis and recommend a data plan on monthly basis.

3. Proposed Method

The recommendation of interesting result using Mobile Internet uses various process such as data collection, data pre-processing, feature extraction, classification etc. The proposed architecture shows the functionality of every module for useful recommendation.

The proposed model (Architectural framework) showed by below stepwise**:**

3.1.1 Data collection using API

a. Data collection is the process of gathering information from all relevant sources to find answers to research questions, test hypotheses and evaluate results. Therefore, we first extract the data log from the API.
b.

3.1.2. Data Processing and Cleaning

a. One Data pre-processing is a data extraction technique that consists of converting raw data into a comprehensible format. Data pre-processing is a reliable way to solve such problems. Data pre-processing prepares raw data for further processing.

b. Cleaning the data of the housing: clean the data by filling the missing values, standardizing noisy data or solving inconsistencies in the data.
c. Data integration: collect data with different representations to resolve data conflicts.
d. Data conversion: data are normalized, aggregated and promoted.
e. That is, data reduction: this phase aims to provide a simplified representation of data in the data warehouse.

3.1.3 Data Classification (Decision Tree)

a. Data Classification is a technique for classify our data in varies class.
b. In classification we use varies technique like decision tree, Linear Regression etc.
c. In decision tree algorithm we classify our data Layer by layer in one or more classes.
d. Decision Tree algorithm we use feature extraction algorithm to reduce our class size.
e. We divided our Data sets in two classes primary and secondary.
f. Audio/Browsing/Message all are primary Data sets.

g. Secondary data source is video streaming. We can reduce and increases to change streaming quality.
h. We design our data classification tree on the basis of used bandwidth. Calling Streaming consumption of data compare to streaming data used different in a single time interval.
i. If a user surfing one hour then they consumed Average 15 Mb/Hour Data. Surfing Consumption depends on user surfing web pages. We can't reduce consumption of data in surfing.
j. If a user streaming, they consumed too high amount of data compare to another activity. If user streaming then reduces and increases data consumption help video quality manipulation

3.1.4 Data Classification Formula

Total consumed data = $\sum$sum of all data=D_{Total}

Call Use data= $\sum$ Call =D_{call}

Audio data= $\sum$ Audio= D_{Audio}

Text Use data= $\sum$ Text=D_{Text}

Web-browsing Use data= $\sum$Browsing data=$D_{Browsing}$

Primary data=$D_{primary}=\sum D_{call} + D_{Text} + D_{Browsing} + D_{Audio}$

Total Video stream data = $\sum D_{Total} - D_{primary}$.

3.1.4 Data Classification Algorithm

```
Input- Used data
   if (data consumption>20MB/Hour)
      if (data consumption>40MB/Hour)
         if (data consumption>120MB/Hour)
            if (data consumption>240MB/Hour)
               secondary Data
                  else
                  streaming data
```

else

Audio Data

else

Call data

else

Browsing Data

3.2 Feature Extraction

a. SeenTime -This parameter defines total stream time in per day. We divide 24 hours into the five-time units A, B, C, D, E.

b. Last_7Days_Average-This parameter same as Average_seen_Timebut it is average data of the last seven specific day average saw the time.

c. Seen Time in days-This parameter represents the user stream time behaviour. we took 24 hours clock and divide the user seen time eight-unit they are following.

3.3 Data Associativity (A Prior Algorithm)

Association rules

There are so many methods which are used to determine the relation of the items by combining any algorithms. The invention of knowledge or the new the system can compare some techniques for a datum, therefore, the new analysis for the invented knowledge is needed. This research discusses the comparison between market basket analysis by using A priori algorithm and market basket analysis without using an algorithm to generate new knowledge

Support Score

Combined percentage of the two items: for identifying the combination of the item which is fulfil the minimum requirement of support value.

Support value of an item is achieved by using the following formula:

$$S(A) = \frac{\text{Amount of transaction A}}{\text{Total transaction}} \quad \text{.............(1)}$$

Confidence Score

The frequencies of the item Y appear in the transaction which contains X. After all of system of high frequency found, then rules need to be found.

$$\text{Confidence} = \frac{\text{frequency of both parameter (A, B)}}{\text{Total no of frequency A}} \qquad \text{Lift} = \frac{\text{No of support of (A, B)}}{\text{Support(A) X Support(B)}}$$

3.4 A priory Candidate Key

Days: - we choose all seven days as a unique day so we have a total of seven unique days as a different parameter. Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday.

Average_seen_Time: -This parameter defines total stream time in per day. We divide 24 hours into the five-time units A, B, C, D, E.

0 TO 2 Hours= A	6 TO 8 Hours= D
2 TO 4 Hours= B	8 TO 24 Hours= E
4 TO 6 Hours= C	

Last_7Days_Average: -This parameter same as Average_seen_Time but it is average data of the last seven specific day average saw the time**.**

Seen_Time in days:- This parameter represents the user stream time behaviour. we took 24 hours clock and divide the user seen time eight-unit they are following.

0 To 3 = F	12 To 15= J
3 To 6 = G	15 To 18= K
6 To 9 = H	18 To 21= L
9 To 12 = I	21 To 24= M.

3.5 Data Estimation Process

a. Above associativity result show basis of user data best frequently used pattern. We used this pattern estimate daily basis and monthly basis data consumption.
b. We calculate total data consumption on the basis of user previous used best possible frequently used pattern given by A priori algorithm.
c. Similarly, we can estimate user all days in week streamed data needs.
d. We add our primary data set in our estimated secondary data sources to estimated total estimation of user data.
e. Similarly, we calculate all days in a week data and then we then add all thirty days data to find monthly consumption data of user.

4.1 Result Evaluation

In this chapter discuss about the experiment evaluation of this project. After generate the result will evaluate the result on the basis of parameter.

4.1.1 Recommendation Result

Our proposed recommendation system uses the relative overall association score generated using our predefined parameter. All parameter defines according to human nature-based. Us predefine parameters **Day, Average_seen_Time, Last_7Days_Average, Seen_Time in days.** we define to all-time slot in a unit we can increase unit according to our requirement but decreasing of unit size complexity increase. we define all parameter one by one.
Days: - we choose all seven days as a unique day so we have a total of seven unique days as a different parameter.

4.1.2 Collect User Used Previous Data

Here this is the data sheet of user which is shown the user used data report. This Data sheet is used to mining to find new information. We need user previous data list report in basis of date and time. To Collect User used previous data report We need to Login Our ISP provider web services. We used to JIO Internet data Service provider. To collect user Used Previous data follows step….

Using MyJio app:

- **i** Sign in to MyJio
- **ii** From the menu select 'My Statement'
- **iii** Specify the start and end dates
- **iv** Tap on 'View' to generate the statement. You can also download the same in PDF format. After download pdf we can change in other data format word data format help of office 2013. Then we copy our Data log in Excel format.

Using Jio.com:

I. Login in to Jio.com through OTP
II. Click on 'My Statement'
III. Specify the start and end dates, with the date range within 30 days.
IV. Click 'View' to generate the statement. You can also download the same in PDF format. Tap on 'View' to generate the statement. You can also download the same in PDF format. After download pdf we can change in other data format word data format help of office 2013. Then we copy our Data log in Excel format.

Jio DIGITAL LIFE

Mr. Navneet Sahni Jio Number : 8700057914 Statement Date : 03-Oct-2019

Itemized Usage

No.	Start Date & Time	End Date & Time	Destination	Total Usage (MB)	Billed Usage (MB) (A)	Free Usage (MB) (B)	Chargeable Usage (MB) (C = A-B)	Amount (₹)
1.0	Data							
1.1	Usage in India							
1.1.1	LTE							
1	03-MAY-19 23:51:13	04-MAY-19 00:51:14	JIONET	0.030	0.030	0.030	0.000	0.00
2	03-MAY-19 23:51:13	04-MAY-19 00:25:06	JIONET	1.374	1.368	1.368	0.000	0.00
3	03-MAY-19 18:06:09	03-MAY-19 23:51:13	JIONET	758.020	758.028	758.028	0.000	0.00
4	03-MAY-19 15:22:58	03-MAY-19 18:06:08	JIONET	14.359	14.356	14.356	0.000	0.00
5	03-MAY-19 10:22:53	03-MAY-19 15:22:58	JIONET	340.705	340.704	340.704	0.000	0.00
6	03-MAY-19 05:22:48	03-MAY-19 10:22:53	JIONET	116.965	116.954	116.954	0.000	0.00
7	03-MAY-19 00:22:43	03-MAY-19 01:22:44	JIONET	0.044	0.049	0.049	0.000	0.00
8	03-MAY-19 00:22:43	03-MAY-19 00:25:06	JIONET	0.039	0.040	0.040	0.000	0.00
9	02-MAY-19 23:22:42	03-MAY-19 00:22:43	JIONET	121.263	121.260	121.260	0.000	0.00

Figure 4.1: Data Sheet of user in PDF Format

We see in above data Log given following information…
Start Date& Time: It is start time of user start internet services.
End Date & Time: It is End time of user internet services.
Destination: It show the destination of internet.
Billed Usage: It show the used data of user.
Chargeable Usage: It show the extra used according to subscribe plan.
Amount: Its shows charge according previous agreement plan.

We need to convert all data in word and excel format to process all data(information)

And mining all data.

Start Date & Time	End Date & Time	Destination	Total Usage (MB)	Billed Usage (MB) (A)	Free Usage (MB) (B)	Chargeable Usage (MB) (C = A-B)	Amount ( )
03-05-2019 23:51	04-05-2019 00:51	JIONET	0.03	0.03	0.03	0	0
03-05-2019 23:51	04-05-2019 00:25	JIONET	1.374	1.368	1.368	0	0
03-05-2019 18:06	03-05-2019 23:51	JIONET	758.02	758.028	758.028	0	0
03-05-2019 15:22	03-05-2019 18:06	JIONET	14.359	14.356	14.356	0	0
03-05-2019 10:22	03-05-2019 15:22	JIONET	340.705	340.704	340.704	0	0
03-05-2019 05:22	03-05-2019 10:22	JIONET	116.955	116.954	116.954	0	0
03-05-2019 00:22	03-05-2019 01:22	JIONET	0.044	0.049	0.049	0	0
03-05-2019 00:22	03-05-2019 00:25	JIONET	0.039	0.04	0.04	0	0
02-05-2019 23:22	03-05-2019 00:22	JIONET	121.263	121.26	121.26	0	0
02-05-2019 21:22	02-05-2019 23:22	JIONET	234.093	234.102	234.102	0	0
02-05-2019 17:41	02-05-2019 21:22	JIONET	269.674	269.678	269.678	0	0
02-05-2019 12:10	02-05-2019 17:41	JIONET	388.766	388.76	388.76	0	0
02-05-2019 07:10	02-05-2019 12:10	JIONET	354.998	355	355	0	0
02-05-2019 00:58	02-05-2019 02:10	JIONET	249.108	249.102	249.102	0	0
02-05-2019 00:14	02-05-2019 00:25	JIONET	38.578	38.584	38.584	0	0
02-05-2019 00:14	02-05-2019 00:58	JIONET	145.438	145.44	145.44	0	0
01-05-2019 19:51	02-05-2019 00:14	JIONET	563.804	563.799	563.799	0	0
01-05-2019 14:51	01-05-2019 19:51	JIONET	90.435	90.44	90.44	0	0
01-05-2019 13:22	01-05-2019 14:51	JIONET	78.58	78.575	78.575	0	0
01-05-2019 08:22	01-05-2019 13:22	JIONET	339.781	339.786	339.786	0	0
01-05-2019 00:22	01-05-2019 01:22	JIONET	16.069	16.065	16.065	0	0
01-05-2019 00:22	01-05-2019 00:25	JIONET	5.294	5.303	5.303	0	0

Figure 4.2: Data Sheet in Excel Form

We see in above data report some extra information. We need to remove extra information. We need remove Destination, Billed usage, Chargeable Usage and amount.

Start Date & Time	End Date & Time	Total Usage (MB)
03-05-2019 23:51	04-05-2019 00:51	0.03
03-05-2019 23:51	04-05-2019 00:25	1.374
03-05-2019 18:06	03-05-2019 23:51	758.02
03-05-2019 15:22	03-05-2019 18:06	14.359
03-05-2019 10:22	03-05-2019 15:22	340.705
03-05-2019 05:22	03-05-2019 10:22	116.955
03-05-2019 00:22	03-05-2019 01:22	0.044
03-05-2019 00:22	03-05-2019 00:25	0.039
02-05-2019 23:22	03-05-2019 00:22	121.263
02-05-2019 21:22	02-05-2019 23:22	234.093
02-05-2019 17:41	02-05-2019 21:22	269.674
02-05-2019 12:10	02-05-2019 17:41	388.766
02-05-2019 07:10	02-05-2019 12:10	354.998
02-05-2019 00:58	02-05-2019 02:10	249.108
02-05-2019 00:14	02-05-2019 00:25	38.578
02-05-2019 00:14	02-05-2019 00:58	145.438
01-05-2019 19:51	02-05-2019 00:14	563.804
01-05-2019 14:51	01-05-2019 19:51	90.435
01-05-2019 13:22	01-05-2019 14:51	78.58
01-05-2019 08:22	01-05-2019 13:22	339.781
01-05-2019 00:22	01-05-2019 01:22	16.069
01-05-2019 00:22	01-05-2019 00:25	5.294
30-04-2019 19:21	01-05-2019 00:22	310.878

Figure 4.3: Clean Data Sheet

Our Sample Size 30 days Data of a single user. Our Sample Size is 34 GB of total used data.
We see in above it is total use data (Audio, Video, Surfing, and music's). We divide in our data set in two class Primary and Secondary data. Two Classify our data set we used Decision tree.

Start Date & Time	End Date & Time	Total Usage (MB)		
03-05-2019 23:51	04-05-2019 00:51	0.03	Friday	Primary
03-05-2019 23:51	04-05-2019 00:25	1.374	Friday	Primary
03-05-2019 18:06	03-05-2019 23:51	758.02	Friday	Secondry
03-05-2019 15:22	03-05-2019 18:06	14.359	Friday	Primary
03-05-2019 10:22	03-05-2019 15:22	340.705	Friday	Secondry
03-05-2019 05:22	03-05-2019 10:22	116.955	Friday	Primary
03-05-2019 00:22	03-05-2019 01:22	0.044	Friday	Primary
03-05-2019 00:22	03-05-2019 00:25	0.039	Friday	Primary
02-05-2019 23:22	03-05-2019 00:22	121.263	Thursday	Primary
02-05-2019 21:22	02-05-2019 23:22	234.093	Thursday	Secondry
02-05-2019 17:41	02-05-2019 21:22	269.674	Thursday	Secondry
02-05-2019 12:10	02-05-2019 17:41	388.766	Thursday	Secondry
02-05-2019 07:10	02-05-2019 12:10	354.998	Thursday	Secondry
02-05-2019 00:58	02-05-2019 02:10	249.108	Thursday	Secondry
02-05-2019 00:14	02-05-2019 00:25	38.578	Thursday	Primary
02-05-2019 00:14	02-05-2019 00:58	145.438	Thursday	Primary
01-05-2019 19:51	02-05-2019 00:14	563.804	Wednesday	Secondry
01-05-2019 14:51	01-05-2019 19:51	90.435	Wednesday	Primary
01-05-2019 13:22	01-05-2019 14:51	78.58	Wednesday	Primary
01-05-2019 08:22	01-05-2019 13:22	339.781	Wednesday	Secondry
01-05-2019 00:22	01-05-2019 01:22	16.069	Wednesday	Primary
01-05-2019 00:22	01-05-2019 00:25	5.294	Wednesday	Primary
30-04-2019 19:21	01-05-2019 00:22	310.878	Tuesday	Secondry

Figure 4.4: Classified Data Sheet

Primary Data Class: -Primary Data class is Fixed Data Like Audio, Message, Email, Call data they are Fixed consumption because we can't manipulate these consumptions. We see in above main data traffic of streaming data. Total primary consumed data is 6.5GB.

06-04-2019 21:30	06-04-2019 23:30	57.628	Saturday	Primary
06-04-2019 18:19	06-04-2019 21:30	1.249	Saturday	Primary
06-04-2019 18:18	06-04-2019 18:19	0.087	Saturday	Primary
06-04-2019 18:06	06-04-2019 18:18	0.168	Saturday	Primary
06-04-2019 15:40	06-04-2019 18:06	57.645	Saturday	Primary
06-04-2019 15:35	06-04-2019 15:39	0.215	Saturday	Primary
06-04-2019 15:30	06-04-2019 15:35	0.326	Saturday	Primary
06-04-2019 14:06	06-04-2019 15:30	1.671	Saturday	Primary
06-04-2019 09:06	06-04-2019 14:06	58.757	Saturday	Primary
06-04-2019 08:58	06-04-2019 09:06	1.786	Saturday	Primary
06-04-2019 07:00	06-04-2019 07:28	0.434	Saturday	Primary
06-04-2019 00:00	06-04-2019 01:00	0	Saturday	Primary
06-04-2019 00:00	06-04-2019 00:25	0.001	Saturday	Primary
		6504.959		

Figure 4.5: Primary Data Class

Secondary Data Class: - Secondary Data class is streaming data class. Secondary Data class all Streaming data. We manipulate our consumption increase and decrease quality of video. More Than 50% of Data traffic is Streaming data. In our sample size also 80% Streaming Data.

		(MB)		
03-05-2019 18:06	03-05-2019 23:51	758.02	Friday	Secondry
03-05-2019 10:22	03-05-2019 15:22	340.705	Friday	Secondry
02-05-2019 21:22	02-05-2019 23:22	234.093	Thursday	Secondry
02-05-2019 17:41	02-05-2019 21:22	269.674	Thursday	Secondry
02-05-2019 12:10	02-05-2019 17:41	388.766	Thursday	Secondry
02-05-2019 07:10	02-05-2019 12:10	354.998	Thursday	Secondry
02-05-2019 00:58	02-05-2019 02:10	249.108	Thursday	Secondry
01-05-2019 19:51	02-05-2019 00:14	563.804	Thursday	Secondry
01-05-2019 08:22	01-05-2019 13:22	339.781	Wednesday	Secondry
30-04-2019 19:21	01-05-2019 00:22	310.678	Wednesday	Secondry
30-04-2019 14:21	30-04-2019 19:21	310.265	Tuesday	Secondry
30-04-2019 05:59	30-04-2019 10:59	312.924	Tuesday	Secondry
29-04-2019 19:14	29-04-2019 23:59	592.657	Monday	Secondry
29-04-2019 09:54	29-04-2019 14:54	345.699	Monday	Secondry
28-04-2019 21:01	28-04-2019 22:49	203.052	Sunday	Secondry
28-04-2019 16:01	28-04-2019 21:01	438.513	Sunday	Secondry
27-04-2019 19:28	27-04-2019 21:58	176.702	Saturday	Secondry
26-04-2019 20:01	27-04-2019 00:05	537.329	Saturday	Secondry
26-04-2019 17:50	26-04-2019 20:01	706.362	Friday	Secondry
26-04-2019 14:21	26-04-2019 17:50	166.179	Friday	Secondry
26-04-2019 08:48	26-04-2019 14:11	202.846	Friday	Secondry
25-04-2019 22:04	26-04-2019 00:02	319.975	Thursday	Secondry
25-04-2019 19:05	25-04-2019 22:04	288.65	Thursday	Secondry
25-04-2019 14:05	25-04-2019 19:05	367.582	Thursday	Secondry

Figure 4.6: Secondary Data Class

After Classification of data we need Add some parameters (Like days, Seen Time, Streaming Time). We can add days help of Python programming Language Method findDay(Date).

findDay(Date): Take input as date in format "dd mm yyyy" and give output days of week.

We calculate total streaming time and add streaming video seen time and add according to our predefine parameters and fill all column.

Start Date & Time	End Date & Time	Total Usage (MB)	Days	Data Class	Streaming Time	Average Streaming
03-05-2019 10:22	03-05-2019 15:22	340.705	Friday	Secondry	C	J,L,M
02-05-2019 00:58	02-05-2019 02:10	249.108	Thursday	Secondry	D	F,L,I,M
01-05-2019 08:22	01-05-2019 13:22	339.781	Wednesday	Secondry	B	I,L
30-04-2019 05:59	30-04-2019 10:59	312.924	Tuesday	Secondry	A	H,K,M
29-04-2019 19:14	29-04-2019 23:59	592.657	Monday	Secondry	A	M,H,K
29-04-2019 09:54	29-04-2019 14:54	345.699	Monday	Secondry	B	J,M
28-04-2019 16:01	28-04-2019 21:01	438.513	Sunday	Secondry	B	k,L
27-04-2019 19:28	27-04-2019 21:58	176.702	Saturday	Secondry	B	L
26-04-2019 08:48	26-04-2019 14:11	202.846	Friday	Secondry	D	I,J,L,M
25-04-2019 08:52	25-04-2019 14:01	375.114	Thursday	Secondry	B	I,J,L,M
24-04-2019 00:34	24-04-2019 01:34	183.809	Wednesday	Secondry	D	F,I,K,L
23-04-2019 08:55	23-04-2019 13:55	322.005	Tuesday	Secondry	B	I,K,L,M
22-04-2019 07:56	22-04-2019 12:56	287.436	Monday	Secondry	C	H,J,K,M
21-04-2019 14:29	21-04-2019 19:55	838.034	Sunday	Secondry	D	K,L,M
20-04-2019 14:41	20-04-2019 20:20	774.49	Saturday	Secondry	D	K,M
19-04-2019 12:49	19-04-2019 17:49	271.259	Friday	Secondry	B	J,K,M
18-04-2019 06:57	18-04-2019 11:34	183.094	Thursday	Secondry	D	I,J,K,L
17-04-2019 05:32	17-04-2019 09:37	202.267	Wednesday	Secondry	C	H,J,L,M
16-04-2019 06:49	16-04-2019 12:35	726.49	Tuesday	Secondry	B	H,I,J,K,M
15-04-2019 08:08	15-04-2019 14:03	362.289	Monday	Secondry	A	J
14-04-2019 06:58	14-04-2019 11:58	313.942	Sunday	Secondry	D	H,J,L,M
13-04-2019 15:26	13-04-2019 18:17	401.724	Saturday	Secondry	C	K,L
11-04-2019 16:01	11-04-2019 21:01	263.153	Thursday	Secondry	A	L

Figure 4.7: Secondary Data Class With parameters

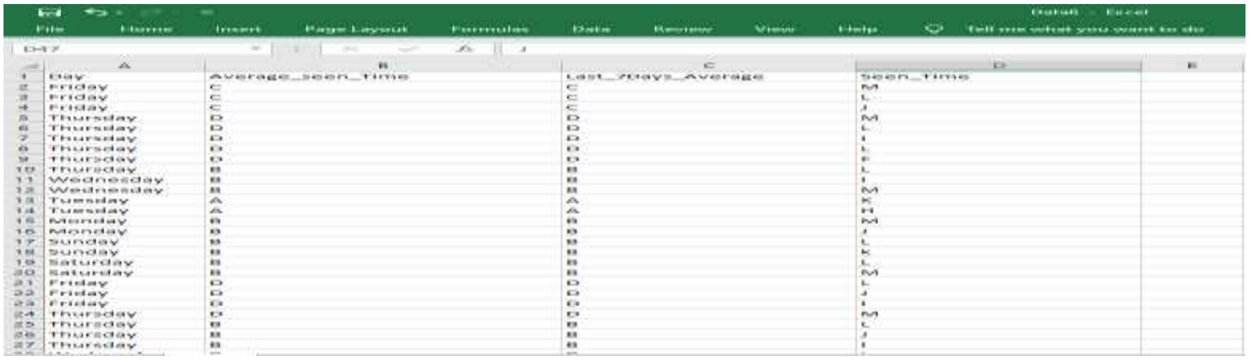


Figure 4.8: Secondary Data Class Input

4.1.3 Simulation

For this work to be done successfully we have used Anconda3 Notebook and MySQL Server.

B. Recommendation of Associativity

Here we have a datasheet, datasheet record from the user data stream of every day with use predefine a parameter. We mining the data sheet information and finding the association score of each and every parameter.

1) Before Recommendation

Here below mention the input window this is displayed. The display data-log they show the data used by the user every day for streaming. We used the four-parameter to describe the user activity those are most influence parameter for the user activity. Those four parameters are Day, Average_seen_Time, Last_7Days_Average, Seen_Time in days.

2) Take some data log as below

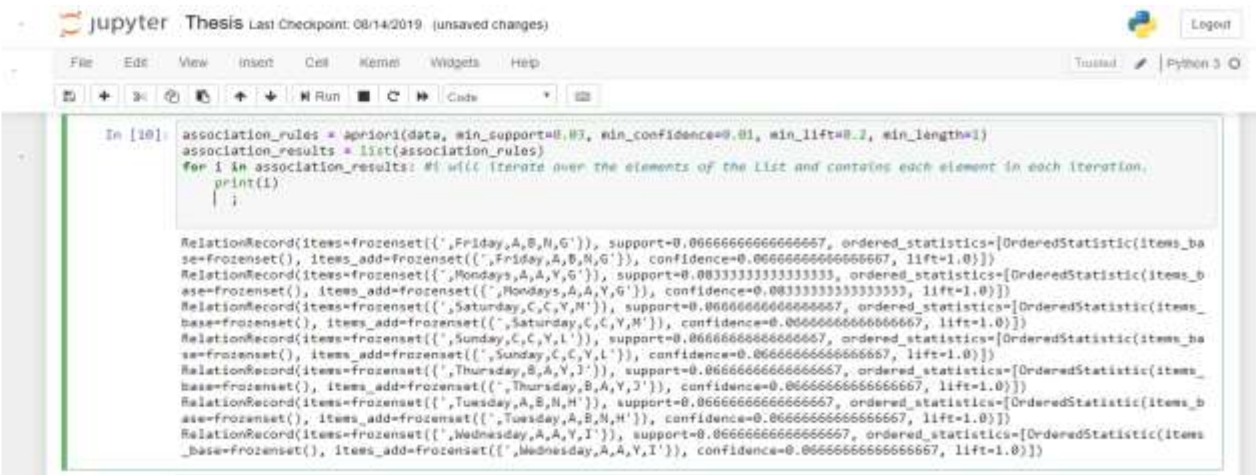
Show the initial screen of proposed work after execution of code this input screen is appearing.

Figure 4.9: Input Window Screen

Above image is image of data to take as input. This input data is data after mining and fill according to our predefine parameter. We analysis these data and finding association we apply association rule in all these data.

3) Analyze & Recommend the associativity relation

After taking the data as input and apply basket analysis (A priori algorithm) or find the all associativity with possible with his lift and confidence score.




```
association_rules = apriori(data, min_support=0.03, min_confidence=0.01, min_lift=0.2, min_length=1)
association_results = list(association_rules)
for i in association_results: #i will iterate over the elements of the list and contains each element in each iteration.
    print(i)

RelationRecord(items=frozenset({',Friday,A,B,N,G'}), support=0.06666666666666667, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({',Friday,A,B,N,G'}), confidence=0.06666666666666667, lift=1.0)])
RelationRecord(items=frozenset({',Mondays,A,A,Y,G'}), support=0.08333333333333333, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({',Mondays,A,A,Y,G'}), confidence=0.08333333333333333, lift=1.0)])
RelationRecord(items=frozenset({',Saturday,C,C,Y,M'}), support=0.06666666666666667, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({',Saturday,C,C,Y,M'}), confidence=0.06666666666666667, lift=1.0)])
RelationRecord(items=frozenset({',Sunday,C,C,Y,L'}), support=0.06666666666666667, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({',Sunday,C,C,Y,L'}), confidence=0.06666666666666667, lift=1.0)])
RelationRecord(items=frozenset({',Thursday,B,A,Y,J'}), support=0.06666666666666667, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({',Thursday,B,A,Y,J'}), confidence=0.06666666666666667, lift=1.0)])
RelationRecord(items=frozenset({',Tuesday,A,B,N,H'}), support=0.06666666666666667, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({',Tuesday,A,B,N,H'}), confidence=0.06666666666666667, lift=1.0)])
RelationRecord(items=frozenset({',Wednesday,A,A,Y,I'}), support=0.06666666666666667, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({',Wednesday,A,A,Y,I'}), confidence=0.06666666666666667, lift=1.0)])
```

Figure 4.10: Output processing Window Screen with Associativity score

We calculate all associativity sets supports score and confidence score. We see in above result for a week days we have multiple associativity sets. We choose best support and confidence score sets.

Associativity Set	Support	Confidence	Lift
Friday, B, B, J	0.0135	0.0135	1.0
Friday, B, B, K	0.0135	0.0135	1.0
Friday, C, C, J	0.0270	0.0270	1.0
Friday, D, D, L	0.0135	0.0135	1.0
Friday, D, D, J	0.0135	0.0135	1.0

Figure 4.11: Output Table with Associativity score

We see in above table Friday associativity result. We have five set with each confidence and support score. We choose C, C, J set for Friday days. It means user Friday streaming chance 6 hour from 12pm to 6 pm more compare to other. Similarly, we can find all seven days best associate result.

Associativity Set	Support	Confidence	Lift
Friday, B, B, (J, k)	0.0270	0.0270	1.0
Saturday, C, C, (K, L, M)	0.0135	0.0135	1.0
Sunday, D, D, (L, M)	0.0270	0.0270	1.0
Monday, C, C, (L, M)	0.0135	0.0135	1.0
Tuesdays, A, A, (H)	0.0270	0.0270	1.0
Wednesday, C, D, (L, M)	0.0135	0.0135	1.0
Thursday, D, D, (L, K, M)	0.0405	0.0405	1.0

Figure 4.11: Output Window Screen with Associativity score full week

4) **After Recommendation**

Here this is our output window of associativity. Here the recommended data are generated in the form of association, confidence and lift score. According to associativity score we calculate the approximate data need for a user.

Data Estimation:

Let assume we calculate data from 6 April 2019 to 3 May 2019.

Days	**Streaming Time (Hour)**
Monday	6 Hours
Tuesday	2 Hours
Wednesday	8 Hours
Thursday	8 Hours
Friday	4 Hours
Saturday	6 Hours
Sunday	8 Hours
Total Streamed	**36 Hours**

C. **4.2 Analysis of Result**

1) **7.2.1 Support**

A Combined percentage of both two-item and the total amount of transaction. Support is the ratio of the frequency both unique candidate and total frequency of transaction. In our proposed method the candidate is our parameters unit. we divide our parameter in a different size we can change our parameter to more accurate result but it also increases the complexity of our method due consider this problem we define our parameter in the medium-size unit. it gives us to approximate accurate result and also decrease our complexity.

$$\text{Support}= \frac{\text{frequency of parameter (Mondays, A, A, G)}}{\text{Total no of transaction}}$$

Support Score (0.0405): It means our four parameter repeats in eights time in hounded time. we can estimate of user consumption according to this parameter because it is most frequent set of user data consumption set.

4.2.2 Confidence

A Combined percentage of both two-item (A->B) frequency and the total no of the transaction of A. Support is the ratio of the frequency both unique candidates combined and total no of the frequency of transaction of A. In our proposed method the candidate parameters divide in size of the unit. we divide our parameter in a different size we can change our parameter to more accurate result but it also increases the complexity of our method due consider this problem we define our parameter in the medium-size unit. it gives us to approximate accurate result and also decrease our complexity. In our case we take confidence zero because our both associativity parameter is transaction is independent.

$$\text{Confidence} = \frac{\text{frequency of parameter (Mondays, A, A, G)}}{\text{Total no of frequency A}}$$

Confidence (0.0405) = In our case support score and confidence order both are same. Confidence score show transaction order in our data order is same so always equal to supports score.

4.2.3Lift

Lift known as ratio of support and multiplication of both candidate frequency; is the number of parameters we want to join together. In our case we chose all four parameters used to find the value of lift in our proposed method all four parameters are important parameter for our result
So, we choose all four parameters to lift and our lift value is four.

$$\text{Lift} = \frac{\text{No of support of (Mondays, A, A, G)}}{\text{Support (Monday) X Support(A)X Support(G)}}$$

4.3 Comparative Performance Evaluation

We have tested our approach using different datasets in multiple languages to confirm the efficiency of the proposed system.

Existing System

Internet Service Provider: -Internet Service Provider reminder system only calculates usage data help of which bandwidth use system. They calculate user data and simply check his plan. When user data usage crosses the limit of his predefine percentage value, they send a remainder system message to user about his use data report. Some ISP provider after cross data limit they reduce the user bandwidth speed help capping system and Some ISP only send system generate message.

All ISP providers after reach of data limit automatically decrease bandwidth and they started base plan according to deal. They did not suggest about how to overcome the data usage. They did not provide any user specific recommendation.

Internet Use Trackers: -Internet use trackers only show the usage data report. These types of tools only show the use of data with use type of service (Like Browsing, Calling, Streaming, etc) and use time. Those are also shown total use data in a period. These are not any recommendation about our total data consumption in the future. They only show the past use of data report.

Proposed System

In our Proposed System, we have taken real-time data from the user. Data are in the form of user data log. Here we took two some datasheet and apply our method to find the association result and we have to check our best association to use the value of support because in our case A->B and B->A equal. So, in our case confidence and support value is same. So, we can choose any one of the values to check best associativity result.

Associativity Set	Support	Confidence	Lift
Friday, B, B, (J, k)	0.0270	0.0270	1.0
Saturday, C, C, (K, L, M)	0.0135	0.0135	1.0
Sunday, D, D, (L, M)	0.0270	0.0270	1.0
Monday, C, C, (L, M)	0.0135	0.0135	1.0
Tuesdays, A, A, (H)	0.0270	0.0270	1.0
Wednesday, C, D, (L, M)	0.0135	0.0135	1.0
Thursday, D, D, (L, K, M)	0.0405	0.0405	1.0

Table 7.12: Proposed System Evaluation Result

Table 4.5 shows the results of our experiments. The table shows Associativity, support, confidence and lift for our proposed system data calculation.

4.4 Data Estimation

First Parameter (Monday): These results represent the associativity result of Monday. This result best associativity result of Mondays.

2nd Parameter and 3rd Parameter (A, A): 2nd parameter represent total streaming time in a day's data. A means users Mondays streaming time is 2 hours it means we need 2 hours streaming data according to user associativity result. 3rd parameter represent the average streaming data of Mondays also 2 hours.

4th Parameter: 4th parameter represent the seen time of user in every day. G means data user seen time 3 to 6 hours.

Data Estimation: Let assume we calculate data from 6 April 2019 to 3 May 2019.

Streaming in 28 Days-

36 X 4 (Week)= 108 Hours

Video Quality	Consumption/Hour	Total Secondary data
144p	80MB	108 X 80
240p	200MB	108 X 200
360p	300MB	108 X 300
480p	500MB	108 X 500
720p	1.5GB	108 X 1.5
1080p	3 GB	108 X 3.0
1440p	5.5GB	108 X 5.5
2160p	14GB	108 X 14.0
4320p	54GB	108 X 54.0

Primary data consumption (28 Days) =6.5 Gb

Day wise= (6.5X1024)/28 Mb

=240Mb

Days wise –

Friday streaming time- 4 Hour (Table 7.3)

Secondary Data – 4(Hour)X 300Mb(360p/Hour)

= 1200Mb

Total Data= 1200Mb+ Primary data

=1200+240=1440MB/Hour

Week Wise

Whole week streaming time- 36 Hour (Table 7.3)

Secondary Data – 36(Hour)X 300Mb(360p/Hour)

= 10800Mb=10.6 Gb

Total Data= 10800Mb=10.6 Gb+ Primary data

28 days

Primary Data = 6.5 GB

Secondary Data(min)=25 GB

Total Data Need (Estimated)= Primary + secondary

= 25 GB + 6.5 GB=31.5 GB

Actual Consume: 34.5 GB

Accuracy= 91.34 percentage.

5. CONCLUSION AND FUTURE WORK

5.1 Conclusion

- In this project, various strategies are discussed regarding data consumption leaning like a percentage of data consumption differently, user-specific features, user behavior for the appropriate prediction of the data estimate.
- Here it provides a view related to several features as a comparison using user's streaming behavior, streaming time, average streaming.
- The different approaches are discussed for analysis like selecting data. The prediction could be made by comparing the number of similarities found according to user streaming behavior.
- Our method is important for both ISP provider and user both ISP estimate the approximate consumption of data calculate every day and use of this they suggest a plan to the user and also use this data for capping.
- it is also working for use, in the same way, to purchase a better plan and reduce bandwidth to fulfil the requirement of the user.

5.2Future Work

- The Future work of this paper will describe Tracking how the behavior of user changes over time we would need to periodically classify the unlabeled parameter.
- As future work, we are planning a deep sensitivity analysis to investigate whether human behavior, user preference and dataset characteristics shape parameters α, β, and γ. We will also include some improvements parameters to the recommendation process. (e.g., person hobby, persons favorite stream program, profession, etc) and we change the divide unit of the parameter.
- In future add some parameter and decrease parameter to increase the accuracy to the proposed method.

References

[1] Christensson, Per. "Data Definition." TechTerms. (2006). Accessed Oct 14, 2019. https://techterms.com/definition/data.

[2] Pandey, Navadha. 2019. "India'S Data Usage Per Smartphone Highest In World At 9.8GB/Month: Report". Https://Www.Livemint.Com. https://www.livemint.com/industry/telecom/india-s-data-usage-per-smartphone-highest-in-world-at-9-8gb-per-month-report-1560936943979.html.

[3] "What Units Of Measurement Are Used For Data Storage?". 2019. Techterms.Com. https://techterms.com/help/data_storage_units_of_measurement.

[4] "Future Mobile Data Usage and Traffic Growth." Ericsson.com, October 29, 2019. https://www.ericsson.com/en/mobility-report/future-mobile-data-usage-and-traffic-growth.

[5] Sen, S., Joe-Wong, C., Ha, S., & Chiang, M. (2012). Incentivizing time-shifting of data: a survey of time-dependent pricing for internet access. IEEE Communications Magazine, 50(11), 91-99. doi: 10.1109/mcom.2012.6353688.

[6] ThirumalaiNambi, E., P.Sakthivel, P. and Krishna, T. (2014). Survey of Evolution Analytical Report on Upgradation Technology from 1G to 5G. International Journal of Computer Applications, 106(18), pp.4-10.

[7] Rani, E. and Juneja, V. (2018). Spectrum Ladder of Mobile Communication Technologies: 1G to 5G. JIMS8I � International Journal of Information Communication and Computing Technology, 6(1), p.343.

[8] B., "Mobile Internet in the news", In Proceedings of the Seventh International Conference on Language Resources and Evaluation LREC, Vol.10,2013.

[9] Andrius, M., Dell, Z., Mark, L., "Combining lexicon and learning based approaches for concept-level Mobile Internet", In WISDOM' Beijing, China, pp.12, 2012

[10] Ricci, F.; Rokach, L.; et al. Recommender Systems Handbook. New York, NY, USA: Springer-Verlag New York, Inc., first edition, 2010, ISBN 0387858199, 9780387858197.
[11] Nagarnaik, P.; Thomas, A. Survey on recommendation system methods. In 2015 2nd International Conference on Electronics and Communication Systems (ICECS), Feb 2015, pp. 1603–1608, doi:10.1109/ ECS.2015.7124857.
[12] Bill Stewart. (2000). IPTO - Information Processing Techniques Office. Retrieved 1 September 2019, fromhttps://www.livinginternet.com/i/ii_ipto.htm
[13] Pazzani, M. and Billsus, D. (n.d.). Content-Based Recommendation Systems. The Adaptive Web, pp.325-341.
[14] Kwak, H., Lee, C., HosungPark, Moon, S.: What is Twitter, a social network or a news media? In: The 19th International Conference on World Wide Web (2010)
[15] ARMOR - Adjusting Repair and Media Scaling with Operations Research for Streaming Video. (2019).Retrieved1 September 2019, from https://digitalcommons.wpi.edu/cgi/viewcontent.cgi?article=1257&context=etd-dissertations.
[16] A Survey Paper on Frequent Itemset Mining Methods and Techniques. (2015). International Journal Of Science And Research (IJSR), 4(12), 545-551. doi: 10.21275/v4i12.nov151884.
[17] Pandas.org. (2019). Pandas [Ebook]. Retrieved from https://pandas.pydata.org/pandas-
a. docs/stable/pandas.pdf
[18] Inmon, W. (2011). Building the data warehouse. 3rd ed. Indianapolis: Wiley, p.35.
[19] Docs.anaconda.com. (2019). FAQs — Anaconda 2.0 documentation. [online] Available at: https://docs.anaconda.com/anaconda-cloud/faq/#what-is-anaconda-inc [Accessed 11 Sep. 2019].
[20] Sun, Y. and Wu, D. (2009). Feature extraction through local learning. Statistical Analysis and Data Mining, 2(1), pp.34-47.
[21] Data Mining Apriori Algorithm. (2018). Presentation.
[22] GISM. (2019). The Mobile Internet Phenomena Report. World: Sandvine. Retrieved from https://www.sandvine.com/hubfs/downloads/phenomena/2019-mobile-phenomena-report.pdf
[23] SURVEY ON APRIORI ALGORITHM FOR MARKET ANALYSIS. (2017). International Journal Of Advance Engineering And Research Development, 4(02). doi: 10.21090/ijaerd.27383.
[24] Khan, A., Jamwal, S., &Sepehri, M. (2010). Applying Data Mining to Customer Churn Prediction in an Internet Service Provider. International Journal Of Computer Applications, 9(7), 8-14. doi: 10.5120/1400-1889
[25] atis. (2012). Glossary – ATIS Telecom Glossary. Retrieved 1 September 2019, from http://www.atis.org/glossary/definition.aspx?id=6555
[26] Bill Stewart. (2000). IPTO - Information Processing Techniques Office. Retrieved 1 September 2019, fromhttps://www.livinginternet.com/i/ii_ipto.htm
[27] What is the difference between a LAN, a MAN, and a WAN?. (2011). Retrieved 1 September 2019, from https://kb.iu.edu/d/agki
[28] Kim, H., Chan, H., & Gupta, S. (2007). Value-based Adoption of Mobile Internet: An empirical investigation. Decision Support Systems, 43(1), 111-126. doi: 10.1016/j.dss.2005.05.009.
[29] Sen, S., Joe-Wong, C., Ha, S., & Chiang, M. (2012). Incentivizing time-shifting of data: a survey of time-dependent pricing for internet access. IEEE Communications Magazine, 50(11), 91-99. doi: 10.1109/mcom.2012.6353688.
[30] What units of measurement are used for data storage?. (2012). Retrieved 6 August 2019, from https://techterms.com/help/data_storage_units_of_measurement.
[31] Petrie, C. (2016). The Proper Use of the Internet: Digital Private Property. IEEE Internet Computing, 20(2), pp.92-94.
[32] Koehler, N. (2013). Healthcare professionals' use of mobile phones and the internet in clinical practice. Journal of Mobile Technology in Medicine, 2(1), pp.3-13.
[33] Lenhart, K. (2019). Social Media & Mobile Internet Use among Teens and Young Adults. Millennials.. [online] Eric.ed.gov. Available at: https://eric.ed.gov/?id=ED525056 [Accessed 1 Sep. 2019].
[34] ThirumalaiNambi, E., P.Sakthivel, P. and Krishna, T. (2014). Survey of Evolution Analytical Report on Upgradation Technology from 1G to 5G. International Journal of Computer Applications, 106(18), pp.4-10.
[35] Simon, Herbert A. "Why should machines learn?." In Machine learning, pp. 25-37. Morgan Kaufmann, 1983.
[36] Ayodele, Taiwo Oladipupo. "Types of machine learning algorithms." In New advances in machine learning. IntechOpen, 2010

[37] Kotsiantis, Sotiris B., I. Zaharakis, and P. Pintelas. "Supervised machine learning: A review of classification techniques." Emerging artificial intelligence applications in computer engineering 160 (2007): 3-24.

[38] Khanum, Memoona, Tahira Mahboob, Warda Imtiaz, Humaraia Abdul Ghafoor, and Rabeea Sehar. "A survey on unsupervised machine learning algorithms for automation, classification and maintenance." International Journal of Computer Applications 119, no. 13 (2015).

[39] Chapelle, Olivier, Bernhard Scholkopf, and Alexander Zien. "Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]." IEEE Transactions on Neural Networks 20, no. 3 (2009): 542-542.

[40] Sutton, Richard S., and Andrew G. Barto. Introduction to reinforcement learning. Vol. 2, no. 4. Cambridge: MIT press, 1998.

[41] "Types Of Machine Learning Algorithms | En.Proft.Me". 2019. En.Proft.Me. https://en.proft.me/2015/12/24/types-machine-learning-algorithms/.

[42] "10 Game-Changing Machine Learning Examples". 2019. Analytics Khabar. https://www.analyticskhabar.com/10-game-changing-machine-learning-examples/.