Users On The Move: On Relationships Between QoE Ratings, Data Volumes and Intentions to Churn

Markus Fiedler[1], Katrien De Moor[2]
[1]Blekinge Institute of Technology
Karlskrona/Karlshamn, Sweden
markus.fiedler@bth.se
katrien.demoor@ntnu.no

Hemanth Ravuri[1], Prithvi Tanneedi[1], Mounika Chandiri[1]
[2]Norwegian University of Science and Technology
Dept. of Information Security & Communication Technology
7020 Trondheim, Norway
{hera15,nata15,moch15}@student.bth.se

Abstract— For a long time, the risk of customer churn, i.e. to leave an operator, has been used as argument in favor of Quality or Experience (QoE) research. However, the understanding of how churn behavior and QoE are related is still limited. This is problematic, as customer retention and churn prediction have grown in importance in face of ever-growing competition on the telecom market. The work presented in this paper aims to make a contribution in this respect, by exploring the relationships between QoE ratings, data volumes and churn risk through a longitudinal user study. Using an Experience Sampling Method-inspired approach we have been collecting weekly feedback on experienced quality, annoyance and intentions to churn from 22 users for up to eight weeks. Additionally, measurements of weekly used data volumes were collected. We observed churning behavior for 3 out of 22 participants and analyze the rating and data usage profiles of churners against non-churners. Furthermore, we investigate correlations of ratings and volumes, and find that "annoyed churners surf less". Our findings point out warning signals for potential user churn as well as promising directions for future studies.

Keywords— Quality of Experience, user ratings, customer churn, data volumes

I. INTRODUCTION

With the upsurge of various services and applications on the Internet, sales of data volume (in Sweden called "surf") became a major source of incomes for Mobile Internet Service Providers (MSPs). Given the fierce competition between them, MSPs need to fight for their piece of the cake and are forced to be proactive in order to increase their users' engagement and the quality of their experiences (QoE) with these very services. Yet, enabling user delight is not necessarily the (only) goal in itself [1]. From the MSP point of view, customer loyalty and increased usage volumes, entailing increased revenues, are highly important.

Indeed, there have been indications that "happy users surf more" [2], implying that good QoE seems to correlate positively with how intensively the service is used, which then may turn into revenue for the corresponding MSP. However, customers do not hesitate to *churn*, i.e., to leave an operator if they don't find what they are looking for, or if they find a better offer somewhere else, both of which entail loss of revenue. As a result, building up and maintaining a long-lasting relationship with users/customers is challenging, as customers' loyalty has to be earned nowadays and can show to be very fragile.

Indeed, customer churn has been a dark side of the story in the telecom industry [3] for many years. In the context of ever-growing global competition and empowering consumer protection regulations (see e.g., [4] for an overview of consumers' rights in the telecom industry in Europe), it has become even more important for telecom operators (and MSPs in particular) to actively work to prevent churn and retain customers that are considering to switch to another provider. Corresponding actions may consist in providing specifically tailored offers or compensations. This in turn means that providers need to be able to efficiently identify customers who are at risk of churning. Whereas a lot of work has focused on churn prediction models based on historical data, there is a lack of empirical studies observing individual users and potential churning behavior in real-time, over longer periods of time, and supported by measurements. Such approaches may inspire the development of near-real time strategies for assessing of churn risk and may lead to more efficient (re-)actions in order to prevent churn.

At this very point, our work aims to make a contribution, as it explores the relationships between QoE ratings, data volumes and churn risk through a longitudinal user study. Using an Experience Sampling Method-inspired approach, we have been collecting weekly feedback on experienced quality, annoyance and intentions with respect to churn from 22 users (of which three users churned unexpectedly) for up to eight weeks. Additionally, measurements of weekly used data volume were collected. In this paper, we explore 1) the relations between experienced quality, annoyance, intention to churn and consumed data volumes, 2) potential differences between non-churners and churners and 3) warning signals for user churn based on our findings.

The paper is organized as follows. The next section briefly provides some background to related work on customer churn. In Section 3, we describe the study set-up and data collection procedure in detail. The analyses and results are presented and briefly discussed in Section 4, which is followed by a number of concluding remarks in Section 5.

II. RELATED WORK

Customer churn (also referred to as "customer defection" in the literature) and customer retention have represented a real challenge for many years in several highly competitive service industries, such as the telecommunication, airlines, electrical

power, banking industries [5, 6]. Churn can be studied from different angles (e.g., customer behavior, intentions to churn, motivations and underlying factors, switching costs and loss revenue, churn prediction, customer retention strategies...). This topic-wise diversity is reflected in the wide range of fields studying customer churn (e.g., marketing, service management, psychology, telecommunications, computer science ...). It also has implications for the approaches used to study churn.

In the telecommunications domain for instance, many studies (see e.g., [7-9]) have addressed the issue of churn prediction, in particular by exploiting data mining techniques. Such approaches are valuable, but as the data used for these studies tend to be historical and network-centric rather than user-centric, they entail important limitations. These can be due to the nature of data set [8], issues with data pre-processing and class imbalance [10], feature selection and derived variables [10, 11] etc. Other work has for instance focused on determinants of switching behavior (e.g., advertising, network, external factors, ...) through survey research [6]. Additionally, case-studies focusing on specific countries [6, 12], as well as comparative studies on switching behavior for different types of service industries [13] have indicated that legislative frameworks and cultural differences play an important role.

In the scope of this paper, we are however mainly interested in churn in relation to QoE. Even though many of the above studies are to a certain extent highly relevant for this relation, the link between QoE and churn behavior (in real time, and thus not based on historical data) is to date still under-explored. QoE and service quality are assumed to be important factors influencing churn, but there is no substantial empirical evidence to support that assumption. One underlying reason in this respect may be that, to the best of our knowledge, there are no publicly available longitudinal datasets containing both QoE and churn-related information that could be exploited for this purpose. Secondly, most of the academic studies on QoE are cross-sectional, based on experimental research designs and thus usually taking place in a controlled lab environment. As a result, there is still a lack of empirical studies on QoE characterized by a high ecological validity, investigating QoE from a longitudinal and real-world perspective.

Whereas collecting user feedback only at one moment in time can be meaningful to gain initial insights into churn behavior (e.g., one could ask customers for the occurrence and timing of antecedents and triggers, potential warning signals in their behavior, etc.), such cross-sectional approaches largely rely on introspection and recall, which both introduce a certain bias. In addition, another motivation underlying the study presented here is that the factors that may play a role in triggering churn behavior cannot meaningfully be studied in isolation (e.g., in a controlled lab environment). As a result, we here opted for a longitudinal and real-life study set-up.

We here also extend our focus to the actual consumption of the service (i.e., consumed data volume) in order to explore how QoE, intention to churn and actual service consumption may be related. In [2] it was investigated how and to which extent user session volumes are related to users' Quality of Experience. Various traffic characteristics (measured on an operational network) were correlated with QoE ratings and the main conclusion was that "happy users surf more". We therefore included an indication of the actual consumption of the service in our own study, which is discussed in more detail in the following section.

III. METHODOLOGY

A. General study set-up

Two main types of data were collected for up to 8 weeks: self-reports and used data volumes (for wireless and mobile connectivity, here we focus on the latter). Using a convenience sampling procedure, a total of 22 users (12 female, 10 male) were recruited for the study. In order to avoid bias in the usage patterns due to the study, the participants were instructed to use their (Android-based) Smartphone as they would normally do during the test period. The occurrence of dropout and challenge of keeping participants engaged are known challenges in longitudinal studies. In our study, participants could decide themselves how long they wished to continue. The majority of the users participated for 5 or 6 weeks (6 and 8 users, respectively). Around one third of the participants continued even longer. One user dropped out of the study after four weeks. The recruited participants resided in 3 different countries: 13 in Sweden, 1 in the US and 8 in India. They were linked to in total 5 different mobile Internet service providers (1 in Sweden, 1 in the US and 3 in India). 84% of the participants are between 18-25 years old and only one of them was bound to the service provider by a contract. The other users used pre-paid cards.

B. Collection of self-report data

The collection of the self-report data was based on principles of the Experience Sampling Method (ESM)[14]. ESM is a reliable and valid method that is particularly suitable for investigating users' experiences 'in the wild'. However, it was not a pure ESM study in the sense that no additional contextual information was gathered. When using experience sampling, participants are typically triggered ("sampled") at certain time intervals and asked to e.g., report on current activities, feelings, or provide feedback in a pre-defined format. In our study, participants were invited to provide feedback on a weekly basis. This interval was selected in order to not burden or annoy participants too much, while at the same trying to capture indications of potential churning behavior on a reasonable basis. The following questions were included:

- *Experienced quality*: Overall, how would you evaluate your mobile Internet quality during the past week? (5-point scale ranging from 1 (bad) to 5 (excellent)).

- *Annoyance*: Overall, to which extent did you feel annoyed with your mobile Internet service during the past week? (5-point scale ranging from 1 (not at all annoyed) to 5 (highly annoyed)).

- *Attitude towards churn*: To which extent would you consider changing your mobile service provider due to your experiences during the past week? (5-point scale ranging from 1 (not at all) to 5 (yes, for sure)).

C. Collection of self-report data

MSPs generally provide information on monthly data usage to their customers. In parallel to this work, we evaluated the feasibility of that monthly usage data, but found it to be too coarse for the purpose of this study. The participants were therefore asked to install the Android application called Android Traffic Grapher (ATG) [15] on their smartphone and to share their data usage every week. ATG works similar to other traffic graphers such as MRTG [16] and monitors network traffic counters in the background periodically for both mobile and Wi-Fi interfaces. The sampling interval can be customized; in our study, the data aggregated per week were collected. We opted for this tool as it also distinguishes between mobile and Wi-Fi data unlike other, similar tools provided in most Android devices. The tool doesn't consume storage on the smartphone itself, thereby also reducing the chance of participants may get annoyed with the tool itself and drop out, as in [17].

IV. ANALYSES AND RESULTS
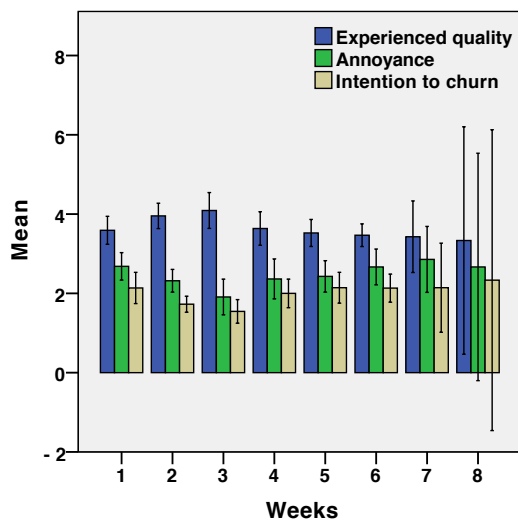
A. Overall observations and correlations

Fig. 1. Average overall ratings with 95% confidence intervals.

From Figure 1 we can observe that the overall ratings for *experienced quality*, *annoyance* and *intention to churn* remained relatively stable over the test period. However, the confidence intervals increased towards the end of the study, indicating a larger spread of the data. We checked for statistical differences in ratings between the weekly feedback moments, but found no evidence in that direction. As we are also interested in the relations between the different indicators that tell us something about the perception and experience from the user perspective, we explored the correlations between the self-report measures and collected data volumes. We therefore calculated the Spearman correlation coefficient r_s and report here only on clear and significant correlations with $|r_s| > .5$.

The results show that lower *experienced quality* ratings go hand in hand with higher *annoyance* ($r_s = -.73$, $p < .01$) and a higher *intention to churn* ($r_s = -.78$, $p < .01$). We can also observe a positive relation between *annoyance* and the *intention to churn* ($r_s = -.74$, $p < .01$), implying that annoyed users are more likely to churn and should therefore be followed with care. A number of other significant correlations were identified, but these were minor and are therefore omitted here. Interestingly, we could not find a clear correlation between the self-report measures and the consumed data volumes. As a result, based on the gathered data and taking into account the limited sample size, we can not support the claim that "happy users surf more" with clear empirical evidence.

As mentioned earlier, three of the 22 participants unexpectedly decided to switch to another provider during the study period. We therefore also checked whether some other trends come to the surface when considering churners vs. non-churners. Overall, the results are in line with the correlations discussed above, with one exception: for the users that churned, the data show a negative correlation between the used *data volume* and the reported *annoyance* level, meaning that annoyed users tend to surf less ($r_s = -.58$, $p < .01$). Given the limited sample and need for more contextual information to interpret this finding, we can not draw any generalizable conclusions. However one possible explanation may be that the annoyed users faced problems that prevented them from surfing more.

B. Differences between non-churners and churners

The boxplots in Figure 2 display the overall distribution of the ratings for the *experienced quality*, *annoyance* and *intention to churn*, as well the overall used *data volume*. We compare non-churners and non-churners.

We can observe that – despite the large range in the ratings for both groups – the reported *experienced quality* was clearly lower for churners than for non-churners. Correspondingly, the users that churned indicated a higher *annoyance* level and a higher *intention to churn*. For the consumed *data volume*, the differences are less pronounced.

In order to verify whether the above observations are also significant from a statistical point of view, we conducted Mann-Whitney tests (a non-parametric equivalent to the independent *t*-test). The obtained results should of course be considered in the light of the limited sample size (for this reason we also report on the effect size *r*).
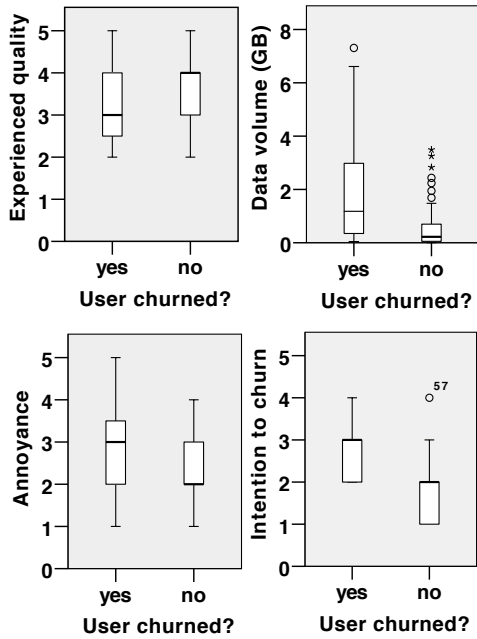
Fig. 2: Box plots displaying the overall ratings for experienced quality, annoyance, intention to churn, as well as the data volumes (GB), per churners vs. non-churners.

Nevertheless, they show a number of interesting trends. To start with, we found that the reported *experienced quality* is significantly higher for non-churners (Median $Mdn = 4$) than for churners ($Mdn = 3$), $U = 791$, $z = -2.31$, $p < .05$, $r = -.19$.
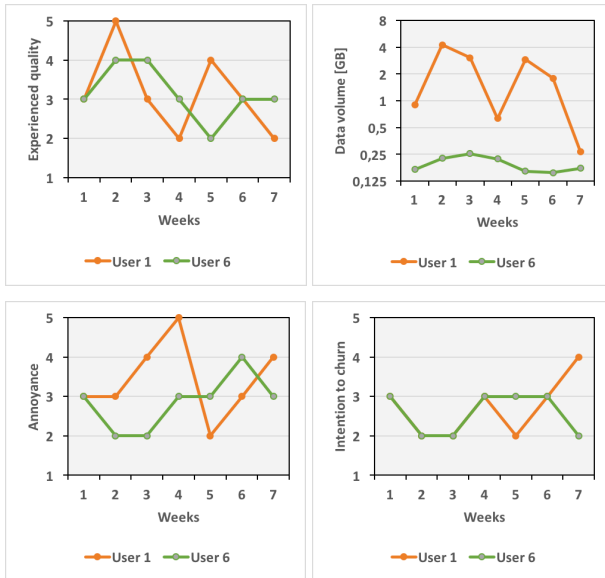

Fig. 3: Comparison of ratings and data volume of user 1 (churner) and user 6 (non-churner).

We now turn our attention towards a comparison of individual ratings of a churner (User 1) with those of a non-churner (User 6) over seven weeks. Figure 3 illustrates the corresponding ratings and data volumes over time.

The *experienced quality* rating shown in Figure 3 does not reveal any systematic difference between the users; the ratings keep passing by each other. However, the churner's quality ratings keep on decreasing during the last three weeks. A similar behavior is seen for the *annoyance* rating; opposite to the quality rating, it keeps on increasing during the last three weeks. The intention to churn even happens to be the same during the first four weeks, while it shows an upwards trend similar to the annoyance. Finally, the *data volume* of the churner varies stronger than that of the non-churner, with drops from weeks 2 to 4 and from weeks 5 to 7, the latter in accordance with the earlier observations.

C. Decision trees

The area of Machine Learning provides powerful tools that help in revealing underlying classification rules for large data sets, aka Big Data [18] despite of the fact that we have data sets of rather limited size, we exploit the exploratory and descriptive power of such methods in order to illustrate potential decisions based on user ratings.

In order to derive such decision trees, we used a Machine Learning classification algorithm (WEKA J48) [19] on four-week sequences of the ratings of three churners and three non-churners in a sliding window fashion, with the two potential outcomes "No churn" and "Churn" at the end of the corresponding period. We define the difference of ratings R a number of i weeks before the potential churn event as

$$\Delta R(-iw) = R(i \text{ weeks before}) - R((i+1) \text{ weeks before})$$

The best-performing decision tree was obtained for the *intention to churn* $R = C$. It is illustrated in Figure 4; the green numbers show the numbers of classifications in the corresponding branches, while the red numbers show the numbers of erroneous classifications (here none).
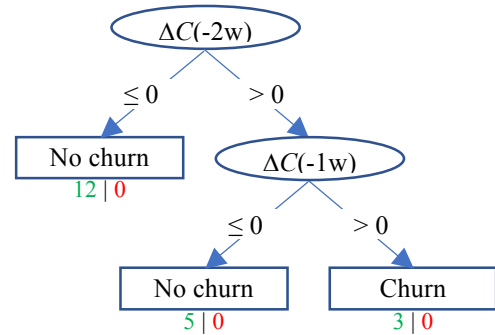

Fig. 4. Decision tree based on recent intentions to churn.

Figure 4 supports the observations of Figure 3 in the way that the churn event is preceded by a two-weeks period of strictly rising intentions to churn. On the other hand, neither experienced quality nor annoyance ratings have shown to be helpful to construct any meaningful decision tree.

In the sequel, we investigate the classification power of the average ratings during seven weeks. Figure 5 shows the decision tree based on the average *intention to churn*.
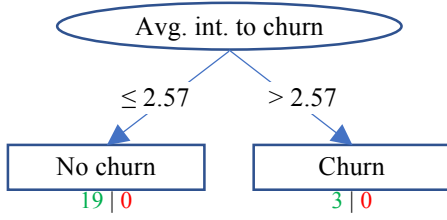
Fig. 5. Decision tree based on average intentions to churn.

Obviously, there exists a distinct threshold (2.57) for the average *intention to churn* as a kind of barometer. In case of the experienced quality ratings, the corresponding threshold 3.14 yields misclassifications. This is not surprising, as the quality ratings as such are much more volatile and affected by contextual factors than the user-expressed right-on-target intention to churn.
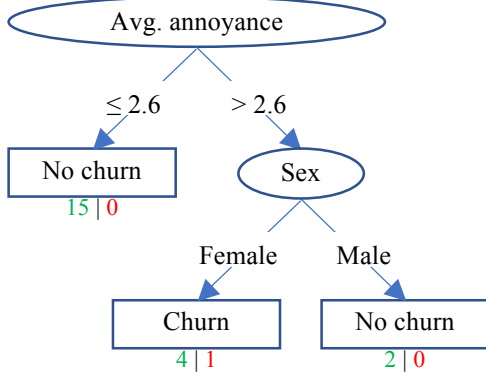

Fig. 6. Decision tree based on average annoyance ratings.

Finally, Figure 6 shows the decision tree based on average *annoyance* ratings. While the threshold for the average annoyance (2.6) is similar to that for the average churn risk rating (2.57), a somewhat surprising result is that, with high average annoyance ratings, gender makes a difference: females are predicted to churn (with one misclassification), while males are predicted to stay with the MSP.

D. Churn criteria

From above results, we may derive a set of churn criteria:
(1) The self-expressed intention to churn has a prime position amongst the potential parameters.
(2) The three most recent weeks are important; sinking quality and/or rising intention-to-churn and annoyance ratings during two weeks indicate a strong risk for churn.
(3) For churners, there is a significant negative correlation between annoyance and volumes ("Annoyed churners surf less").
(4) Non-technical factors such as gender may play a role as well.

V. CONCLUSION AND FUTURE WORK

The starting point for this study was to investigate to which extent participants' experienced quality and annoyance ratings, as well as the related data volumes, are linked to the intention to churn (based on the earlier impression that "happy users surf more"). Furthermore, the approach should be usable as early warning signal for the intention to churn, instead of just performing an offline analysis. To this end, we conducted a longitudinal study collecting data from 22 users for up to eight weeks. We gathered ratings for experienced quality, annoyance and intention to churn, while also collecting the consumed data volumes on a weekly basis. We were lucky to observe three unplanned churns, which allowed us to compare churners with non-churners.

Our findings indicate that the mantra that "happy users surf more" may need to be rephrased to "annoyed churners surf less". Indeed, asking the user for her intentions to churn provided the most reliable churn risk indicators in our study, followed by annoyance ratings, experienced quality ratings and volumes. The results also point to a potential gender difference: female participants tended to be more critical and potentially more consequent (all three churners were female). Due to the limited sample size, we cannot make any strong claims in this direction, but the findings open up interesting questions for follow-up work. Additionally, asking the user too often, too much or in a too obvious way is hardly desirable. More longitudinal work is therefore needed to deepen and extend the exploration of "implicit" early warning strategies, e.g., sinking volumes, which may point at annoyance and may activate a strategy for monitoring the user's intentions more closely.

Finally, it is necessary to consider the motivations for churning and the technical, human and contextual factors that may play a role in this respect more in-depth through follow-up longitudinal studies. For instance, the type of subscription (contract vs. pre-paid) and the monthly cost are likely to have an impact on the actions and "freedom to go" of the user. Similarly, the typical use pattern of a user (e.g., type of services and applications that are most frequently used and how they are used) may bring about certain requirements in terms of the expected service quality, which may in turn affect a user's motivation to stay loyal or to switch to another provider. Deeper insights into this interplay of potential influence factors is therefore essential in order to fully understand whether and under which circumstances QoE and quality-related issues may fuel churn intentions and to develop adequate response strategies in order to prevent churn.

REFERENCES

[1] M. Dixon, K. Freeman, and N. Toman, "Stop Trying to Delight your Customers," *Harvard Business Review,* no. July/August, 2010.

[2] J. Shaikh, M. Fiedler, and D. Collange, "Quality of experience from user and network perspectives," *Annals of telecommunications-annales des telecommunications,* vol. 65, no. 1-2, pp. 47-57, 2010.

[3] W. Yu, D. N. Jutla, and S. C. Sivakumar, "A churn-strategy alignment model for managers in mobile telecom," in *3rd Annual Communication Networks and Services Research Conference (CNSR'05)*, 2005, pp. 48-53.

[4] European Commission. (2015). *Users' rights in the Digital Single Market*. Available: https://ec.europa.eu/digital-single-market/en/users-rights

[5] T. O. Jones and W. I. Sasser, "Why satisfied customers defect," *Harvard Business Review,* vol. 73, no. 6, pp. 1-14, 1995.

[6] S. K. Chadha and N. Bhandari, "Determinants of Customer Switching towards Mobile Number Portability," *Paradigm,* vol. 18, no. 2, pp. 199-219, 2014.

[7] Y. He, Z. He, and D. Zhang, "A Study on Prediction of Customer Churn in Fixed Communication Network Based on Data Mining," in *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery*, 2009, vol. 1, pp. 92-94.

[8] A. Idris, A. Khan, and Y. S. Lee, "Genetic Programming and Adaboosting based churn prediction for Telecom," in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2012, pp. 1328-1332.

[9] L. Bin, S. Peiji, and L. Juan, "Customer Churn Prediction Based on the Decision Tree in Personal Handyphone System Service," in *2007 International Conference on Service Systems and Service Management*, 2007, pp. 1-5.

[10] A. Idris and A. Khan, "Customer churn prediction for telecommunication: Employing various features selection techniques and tree based ensemble classifiers," in *2012 15th International Multitopic Conference (INMIC)*, 2012, pp. 23-27.

[11] M. R. Khan, J. Manoj, A. Singh, and J. Blumenstock, "Behavioral Modeling for Churn Prediction: Early Indicators and Accurate Predictors of Custom Defection and Loyalty," presented at the Proceedings of the 2015 IEEE International Congress on Big Data, 2015.

[12] J. H. Ahn, S. P. Han, and Y.-S. Lee, "Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry," *Telecommunications Policy,* vol. 30, no. 2006, pp. 552-568, 2006.

[13] P. G. Patterson and T. Smith, "A cross-cultural study of switching barriers and propensity to stay with service providers," *Journal of Retailing,* vol. 79, no. 2, pp. 107-120, 2003.

[14] M. Csikszentmihalyi and R. Larson, "Validity and reliability of the experience-sampling method," in *Flow and the foundations of positive psychology*: Springer, 2014, pp. 35-54.

[15] . *Android Traffic Grapher (ATG)* Available: http://android-apk.org/rs.in.luka.android.traffic/

[16] . *MRTG - The Multi Routher Traffic Grapher*. Available: http://www.mrtg.org

[17] S. Ickin, K. Wac, M. Fiedler, L. Janowski, J. H. Hong, and A. K. Dey, "Factors influencing quality of experience of commonly used mobile applications," *IEEE Communications Magazine,* vol. 50, no. 4, pp. 48-56, 2012.

[18] BTH. (accessed 2017, 31-05). *BigData@BTH project homepage* Available: http://www2.bth.se/bloggar/bigdata/

[19] Weka. (accessed 2017, 31-05). Available: http://www.cs.waikato.ac.nz/ml/index.html