Towards Real-Time Customer Satisfaction Prediction Model for Mobile Internet Networks

Ayisat W. Yusuf-Asaju[1,2(✉)], Zulkhairi B. Dahalin[2], and Azman Ta'a[2]

[1] University of Ilorin, Ilorin PMB 1515, Nigeria
ayisatwuraola@gmail.com
[2] Universiti Utara Malaysia, 06010 Sintok, Malaysia
zuldahlim@gmail.com, azman@uum.edu.my

Abstract. Satisfying the customers' service requirements and expectation, especially customer satisfaction had been one of the major challenges faced by the mobile network operators in most telecommunication organizations. This article implemented an analytical customer satisfaction prediction model by employing the mobile internet traffic datasets collected in real-time through the drive test measurement. To this end, the implementation phase has employed machine learning algorithms in the Microsoft Machine Learning R client Server. The results show that previous user's traffic datasets can be used to predict customer satisfaction and identify the root cause of poor customer experience before the complete deterioration of the service performance, which could lead to larger percentage of customer dissatisfaction. The mobile network operators can also use the proposed model to overcome the drawbacks of the conventional subjective method of analysing customer satisfaction.

Keywords: Telecommunication · Quality of experience
Mobile network operators · Traffic datasets · Machine learning algorithms
Mean opinion score

1 Introduction

In mobile telecommunication (telecoms) organizations, satisfaction is an individual judgement after the customer experience with a service. Accordingly, the competitive nature of the telecoms market, have made the Mobile Network Operators (MNOs) focus on both the technical aspect of the services provided to the customers, customer expectations and experiences with the aim of satisfying the customers. Generally, quality of experience (QoE) is used to analyse customer perception based on the network performance. Previous studies have demonstrated QoE as a measure of customer perception in relation to the technical parameters representing the quality of service performance, context, expectations and others factors that can influence the customer perception to determine the diversity of satisfied and dissatisfied customers using a specific service [1]. It should be noted that customer expectation is an ideal standard presented as a service level agreement (SLA), which is an agreement between the customers and the MNOs about the service characteristics provided by the MNOs

F. Saeed et al. (Eds.): IRICT 2018, AISC 843, pp. 95–104, 2019.
https://doi.org/10.1007/978-3-319-99007-1_10

[2], usually monitored by the telecoms regulators. In line with the QoE description, both the expectation and other factors are to be used to analyse the overall perceived QoE for a specific service to determine the mean opinion score (MOS).

The reliable and standard method of measuring the perceived QoE (i.e., MOS) is through the subjective method. However, the subjective method is time consuming, lacks usability in real time, lacks repeatability in real time, and in most cases, the study is usually conducted after the service consumption, which might be too late to take actions to prevent customer dissatisfaction [3]. Thus, there is a need for an analytical forecasting model for telecoms network service operation that can predict subjective measurement of the perceived QoE from the objective measurement in real time or near real time.

MOS is often used to estimate the perceived QoE through the prediction of the subjective measurement from the objective measurement [3, 4]. However, recent studies have demonstrated MOS was not sufficient enough to quantify the diversity of satisfied and dissatified customers, because MOS often eliminates the diversity of the customer assessment while quantifying the customer satisfaction [5]. Accordingly, this article improved and implemented an analytical customer satisfaction (ACSAT) prediction model proposed in the study of Yusuf-Asaju et al. [6] to demonstrate the effectiveness of predicting customer satisfaction with the mobile network realtime dataset. The remainder of this article discusses customer satisfaction issues in telecommunications, mobile internet QoE, ACSAT prediction model, the research methodology, analysis, discussion, and conclusion.

2 Literature Review

2.1 Customer Satisfaction in Telecommunications

Previous literatures have provided several thoughtful interpretations on customer satisfaction. Within this context, previous literature has demonstrated performance-specific expectation and expectancy disconfirmation play significant role in customer satisfaction analysis [7]. However, in different cases, disconfirmation measure has the greatest impact on satisfaction. As a result, Xiao and Boutaba [7] hold the view that there is a substantial basis that can be used to derive an analytical forecasting model for the telecoms network services through the use of theories, models of perceived service quality and customer satisfaction adopted from the marketing and economics literatures. Consequently, the authors proposed an analytical modelling approach termed Customer satisfaction (CSAT) model. The CSAT model consists of three service utilities (Network Quality of Service (QoS), Network Availability, and Customer care), expectation update, perceived utilities, disconfirmation, customer satisfaction, market dynamics, customer behavior and profitability and deployed through a mathematical model.

A recent study enhanced the CSAT model by defining the service utilities elements as four QoE dimension (Network coverage, service availability, QoS support and Price) of customers satisfaction [8]. Service content and security was included to the service utility in the CSAT model [8]. The study holds the view that QoS is closely related to perceive QoE and customer satisfaction model is based on QoS criteria and parameters.

Therefore, this study strongly believed that CSAT model can be modified and improved through the QoE components, QoE measurements, and QoE estimations.

2.2 Mobile Internet Quality of Experience

Qualinet in [1] describes QoE as the "*degree of delight or annoyance of the user of an application or service. It results from the fulfilment of his or her expectations with respect to the utility and/or enjoyment of the application or service in the light of the user's personality and current state.*" Perceived QoE measurements are classified into subjective and objective measurement [3]. The subjective measurements are based on customer perception of the services delivered to the customers, while objective measurement is the means of estimating subjective quality solely from the measurement obtained from the network traffic [3, 4]. The MOS is an opinion score on five-point category-judgement scales and it is mostly used in many applications such as audio, video, and data to estimate the perceived QoE [4]. The MOS is illustrated by allocating values to the scores such as Excellent = 5; Good = 4; Fair = 3; Poor = 2; Bad = 1 [5].

According to Le Callet et al. [1], QoE influence factor is "*any characteristic of a user, system, service, application, or context whose actual state or setting may have influence on the Quality of Experience for the user*". This description implies the QoE Influence factors can be classified into system, context and human influence factors. System QoE Influence Factor comprises of the properties and characteristics of the technical quality of the application [1]. This technical quality consists of the QoS parameters (such as throughput, delay, jitter, and loss). Several studies have analysed perceived QoE by employing the delay, jitter, and loss [4, 9]. However, limited studies have used throughput for the analysis of the perceived QoE. According to Battisti et al. [10], throughput is the actual data transmission speed experienced by the internet users.

Context QoE Influence Factor is any situational property that describes the users' environment [1]. Generally, previous studies often combine context factors with human and system factors without any specific structure or categorization [11]. In mobile network scenario, context factors in this study represents the physical components. Physical component describes the characteristics of location and space [11]. Therefore, the physical component variables used for the implementation of the ACSAT prediction model was location (longitude and latitude) and time of the day.

Human QoE Influence Factor describes any characteristics of users including demographic, socio-economic background, socio-cultural background, physical and mental constitution, and emotional state [1, 11]. Although, human QoE influence factor is an important dimension, human factor has been limitedly studied in most empirical studies, due to the difficulties in assessing the way it influences QoE [11]. Moreso, user demographic is difficult to ascertain in real-time, since the MNO often analysed the user experience based on contextual factors. Therefore, this study considers user expectation as SLA. In addition, this study considers the use of region, which describes the population densities of users in an area. This is an example of socio-cultural. Both expectation and region are used as a variable of the human influence factors for the implementation of the ACSAT prediction model.

2.3 Analytical Customer Satisfaction Prediction Model

The ACSAT prediction model comprises of five different assumptions. The first assumptions is to employ network performance as the underlying component of ACSAT prediction model. Network performance is a principal determinant of customer satisfaction. The second assumption is the use of customer traffic data consisting of the mobile internet usage dataset. This assumption was as a result of the limited use of large scale database comprising of customer behaviour to estimate the percieved QoE [12]. The third assumption is to incorporate the QoE influence factors in the ACSAT prediction model. Therefore, modelling of the perceived QoE, considers all the three perceived QoE dimensions of the influence factors in the ACSAT prediction model, which enables an adequate estimation of the perceived QoE in relation to mobility (such as time and location).

The fourth assumption is the estimation of perceived QoE with respect to expectation as indicated in SLA. Several marketing and perceived QoE literatures reviewed showed that expectation is an influence factor of perceived QoE [1, 12]. Although, the expectation in respect to SLA was used in the CSAT model to model the perceived service quality [7]. Tsiaras and Stiller [13] specifically stated that using expectation as stated in SLA would be adequate for the modelling of perceived QoE. In addition, the CSAT model reversed the expectation to formulate the expectation update [10]. However, Ibarrola et al. [2] stated that web survey is usually built to collect customer requirements and preferences. This paper argues that the arrow has to indicate that expectation can used to estimate the perceived QoE and at the same time the customer behaviour observed while estimating the perceived QoE can be used to update the expectation.

Lastly, the fifth assumption proposed in the ACSAT prediction model is perceived QoE maximization to determine the variation of customers that are satisfied and not satisfied. Therefore, this article used threshold of acceptance in Standard deviation of Opinion Scores (SoS) hypothesis (that is, the rate of perceived quality is either acceptable or not acceptable) [5]. The percentage of Good or bad and percentage of poor or worse (%GoB and %PoW that is: ratio of satisfied and dissatisfied customers) can be defined as GoB (u) = P_u ($U \geq 60$) and PoW(u) = P_u ($U \leq 45$). This depicts categories 1–3 on the ACR MOS can be translated into dissatisfied (PoW(u) = P_u ($U \leq 45$)), while categories 4–5 on the ACR MOS can be translated into satisfied (GoB (u) = P_u ($U \geq 60$)) as explained by Hoßfeld et al. [5], where U is the random variable (throughput variable in this case) for quality ratings. All the five assumptions facilitated the process of determining the distribution of satisfied and unsatisfied customers and the results can be used by the MNOs for allocation of network resources [7, 14]. Figure 1 depicts the ACSAT prediction Model.

3 Methodology

Following the underlying assumptions stated in the ACSAT prediction model and description of the QoE influence factor dimension in the ACSAT prediction model, the following research hypothesis were verified:

Hypothesis 1: Time of the day has significant influence perceived QoE.
Hypothesis 2: Location of the users has significant influence on perceived QoE
Hypothesis 3: QoS parameters (Throughput) has significant influence on perceived QoE.
Hypothesis 4: Region of the users has significant influence on perceived QoE.
Hypothesis 5: Expectation (SLA) has significant influence on perceived QoE.

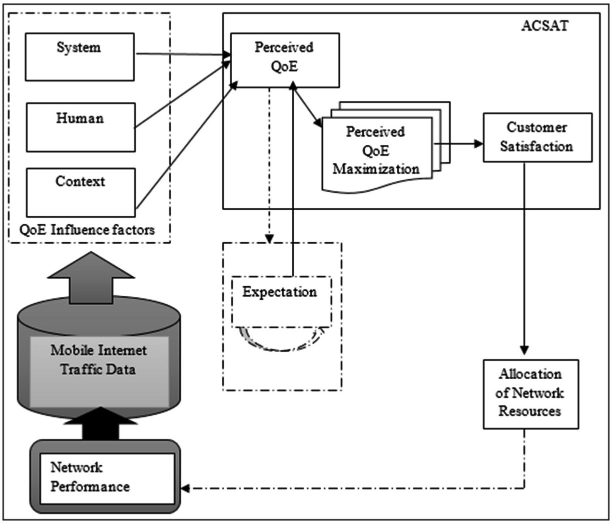

Fig. 1. ACSAT prediction model

3.1 Data Collection

The dataset used in this article was collected in real-time through the drive measurement with the aid of Test Mobile System (TEMS) investigation. TEMS investigation is a software used by the MNOs to collect large amounts of dataset with real-time presentation [15]. TEMS investigation software was used to collect dataset from the real-time measurement of a major MNO in Nigeria at different times of the day for File Transfer Protocol Download (FTP DL) and Hyper-test Transfer Protocol (HTTP) applications because limited studies have evaluated the perceived QoE of these mobile internet applications. The dataset was collected from four different regions (Dense Urban, Urban, Semi-Rural and Rural) in Nigeria. The total observation of the dataset for FTP DL application was 1.8 million while the total dataset of the HTTP application was 1.3 million observations. FTP is a protocol used to facilitate exchange of data between a server and clients(s) and is used as a test due to the specific nature of the data exchange and simulates file download. HTTP simulates general browsing which is the underlying protocol used in WWW and its data exchange.

3.2 Data Preparation

The dataset was collected from the mobile network in real-time as discussed in the data collection section. To make the dataset suitable for the modelling of perceived QoE, the dataset passed through a data preparation process. The dataset was cleaned, integrated, and transformed into a form that was suitable for the modelling phase. In addition, the conceptual labels used were based on the population densities of the regions. Region with the highest population density was represented by 4 (Dense Urban), second highest by 3 (Urban), third highest by 2 (Semi-Rural) and the lowest by 1 (Rural). Similarly, this study mapped the throughput variable (aggregated total application layer throughput used by the users on a particular node) into the Absolute Categorical Rating (ACR) MOS scale, represented as a discreet value (that is 5 = excellent. 4 = good, 3 = fair, 2 = poor, and 1 = bad). Equally, SLA of the QoE influence factor dataset that was used to represent the expectation of the mobile internet subscribers was defined into different categorical interval (5 = Excellent, 4 = good, 3 = fair, 2 = poor, 1 = bad) based on the maximum throughput of 42 Mbps that could be achieved on a node as explained by Ericson [16]. The interpretation was based on the maximum and minimum value of the throughput variable for the ACR scale was achieved through probability mass function of the discrete random variable (throughput). Afterwards, the longitude and latitude variable was combined to represent the single location of the user in one single variable. Similarly, the date and time of the day was also combined to represent a single variable time of the day. The descriptive statistics of maximum and minimum throughput variable interpretation into ACR MOS and SLA score are depicted in Table 1.

Table 1. Descriptive statistics of throughput variable interpretation to ACR MOS and SLA score

FTP DL applications		HTTP applications				
TT variable intervals (Mpbs)	ACR MOS score	TT variable intervals (Mpbs)	ACR MOS score	TT variable (Mpbs)	SLA score	Scale interpretation
0–7.2	1	0–6.046	1	0–8.4	1	Poor
7.3–14.3	2	6.047–12.092	2	8.41–16.8	2	Bad
14.4–21.5	3	12.092–18.138	3	16.81–25.2	3	Fair
21.6–28.7	4	18.139–24.184	4	25.21–33.6	4	Good
28.8–35.8	5	24.185–30.23	5	33.61–42.00	5	Excellent

3.3 Data Modelling

This phase involves the modelling of perceived QoE through an abstract representation of data and its relationship within the dataset. The stage adopts the use of regression task in the form of a predictive technique through the machine learning algorithms (such as multiple regression, decision trees, random forest, decision forest and neural network) in Microsoft Machine Learning R Server Platform. Following the deterministic mathematical (DQX) model in [13] that gives QoE formalization as $QoE \coloneqq f(User, Service, Variable)$. A multiple regression was considered to model the relationship between the MOS and the QoE influence factors (variables) through a linear predictor functions, where unknown model parameters are estimated from the data. In this case, the observation of data instances represents the independent variable (that is, the variables in the QoE influence factors dataset) while the predicted variable is the possible values of dependent variables (perceived QoE) which is the outcome. The predicted perceived QoE was evaluated with root mean squared error (RMSE) [17], which is the most used prediction accuracy technique for numerical measures of QoE in previous studies [4, 9].

4 Results and Discussion

The first analysis conducted was multiple regression analysis between the QoE influence factors and the perceived QoE. The results of the correlation and multiple regression for both FTP DL and HTTP applications are presented in the Tables 2 and 3.

Table 2. Correlation and multiple regression result for FTP DL application

Variables	Pearson correlation	Regression estimates	T-value	P-value
QoE ~ Time of the day	−0.036***	0.005	16.71	<2e−16
QoE ~ Location	0.332***	0.019	48.47	<2e−16
QoE ~ Throughput	0.908***	0.417	744.17	<2e−16
QoE ~ Region	0.010***	−0.291	−38.27	<2e−16
QoE ~ SLA	0.903***	0.555	685.97	<2e−16

***Significant ($p < 0.000$)

Following the results presented in Tables 2 and 3, it can be concluded that all the proposed hypotheses were significant ($p < 0.000$) with R^2 of 0.871 and 0.869 for FTP DL and HTTP applications respectively. The results depict all the stated hypotheses have significant influence on the overall perceived QoE. Specifically, for both applications, H1 was accepted and shows a positive significant influence for MOS in both application, this implies that the MOS increases with the time of the day. This is evident among the mobile internet users whereby, the internet services is usually stable during the late hours of the night (off peak time). H2 was accepted for both application and shows a positive significant influence for MOS in both applications. This implies, MOS increases in irrespective of the user's location. In other words, internet users in different locations are bound to experience different mobile internet experience. H3 was accepted for both application, this interprets the MOS increases with the MOS, the higher the throughput, the higher the mobile internet experience. H4 was accepted for both application with a negative significance influence, this indicates the lower the population density (less congested areas) the higher the mobile internet user experience. Lastly, H5, was accepted and shows a positive significance influence which interprets the expectation of the user increases with the mobile internet experience.

Table 3. Correlation and multiple regression result for HTTP application

Variables	Pearson correlation	Regression estimates	T-value	P-value
QoE ~ Time of the day	0.173***	0.013	36.55	<2e−16
QoE ~ Location	0.033***	0.019	36.82	<2e−16
QoE ~ Throughput	0.915***	0.666	994.93	<2e−16
QoE ~ Region	−0.071***	−0.022	−20.50	<2e−16
QoE ~ SLA	0.855***	0.293	270.02	<2e−16

***Significant ($p < 0.000$)

Overall, the regression results show that all the selected QoE influence factors have significant influence on mobile internet users perceived QoE for both FTP DL and HTTP applications.

The second Analysis conducted was the prediction of the perceived QoE with decision trees, random forest, decision forest and neural network. While the actual MOS of FTP DL application was 1.5018 and actual MOS of HTTP application was 1.6170. The best machine learning accuracy was selected based on the predicted MOS that has the closer value to the actual MOS. The RMSE results for both the FTP DL and HTTP applications are presented in the Tables 4 and 5 below.

Table 4. Perceived QoE modelling accuracy result for FTP download applications

Machine learning algorithms	RMSE	MOS
Multiple linear regression	0.274	1.5025
Decision trees	0.120	1.5024
Random forest	0.118	1.5021
Decision forest	0.072	1.5019
Neural network	0.141	1.4937

Table 5. Perceived QoE modelling accuracy result for HTTP applications

Machine learning algorithms	RMSE	MOS
Multiple linear regression	0.310	1.6162
Decision trees	0.116	1.6172
Random forest	0.127	1.6169
Decision forest	0.091	1.6171
Neural network	0.148	1.6013

The third analysis is the prediction of the customer satisfaction model using logistic regression to determine the diversity of the satisfied and dissatisfied customers. In compliance with the SOS hypothesis discussed in Sect. 2.3, the threshold of %GoB = P (MOS $\geq$ 4) and %PoW = P (MOS $\leq$ 2) was used to filter out the satisfied and dissatisfied quality ratings (accepted or not accepted) from the predicted QoE at Pearson correlation MOS $\sim$ TT equals 0.908 (see Table 2) and Decision forest modelling RMSE was 0.072 (see Table 4) of the FTP DL application. Therefore, a correlation was conducted to ascertain the relationship between the predicted MOS and acceptability variable. Afterwards a binary classification prediction modelling was conducted to analyses the percentage of satisfied and dissatisfied users through logistic regression in R platform. The result of the customer satisfaction analysis of FTP DL application shows a correlation of 0.529 between the perceived QoE (MOSp) and acceptability threshold of the SoS hypothesis. And the binary classification prediction indicates logistic regression have the accuracy of 100% (Specificity = Recall = Sensitivity = Precision = F1 measure = 100%). Following the SoS hypothesis benchmark, the correlation of MOSp and acceptability was 0.529 interprets 97.5% is the %PoW was not Dissatisfied and 2.5% is the %GoB was Satisfied.

For the HTTP application, the correlation of MOS $\sim$ TT equals 0.915 (see Table 3) and Decision forest modelling RMSE = 0.091 (see Table 5) of the HTTP application. And the binary classification prediction indicates logistic regression have the accuracy of 100%. Following the SoS hypothesis benchmark, the correlation of MOSp and Acceptability was 0.637 (at accuracy = Specificity = Recall = Sensitivity = Precision = F1 measure = 100%) interprets 95% is the %PoW = Dissatisfied and 5% is the %GoB = Satisfied. Overall, based on the SoS hypothesis, the percentage of dissatisfied users is 95% and percentage of satisfied users is 5% of the HTTP application. Following the presented results, the MNOs can improve allocation of resources in areas where the perceived QoE is below the acceptable threshold.

5 Conclusion

This article presented an ACSAT prediction model to estimate customer satisfaction of two mobile internet applications using the datasets collected in real-time. The empirical evaluation indicated that customer satisfaction is feasible in real-time to overcome the drawbacks that often occurs through the subjective measurement of customer satisfaction analysis. The analysis showed that the time of the day, location, region, throughput, and expectation have a significant influence on the user perceived QoE. Following the dataset that was used in this study, it was observed that high population density with inadequate network resources can be a cause of the poor mobile internet experience. One way the MNOs can rectify this type of issue is to trace the location and region with very poor MOS and provide proactive measures to increase the network resources that would be efficient for areas with high population density. As seen in the results poor QoE lead into a poor customer satisfaction. Therefore, providing a proactive measures to improve the QoE would reduce the poor rate of mobile internet customer satisfaction.

References

1. Le Callet, P., Möller, S., Perkis, A.: Qualinet White Paper on Definitions of Quality of Experience Output version of the Dagstuhl seminar 1218. 12 June 2012. http://www.qualinet.eu/images/stories/whitepaper_v1.1_dagstuhl_output_corrected.pdf
2. Ibarrola, E., Liberal, F., Ferro, A.: Quality of service management for ISPs: a model and implementation methodology based on the ITU-T recommendation E.802 framework. IEEE Communications Magazine, pp. 146–153 (2010)
3. Tsolkas, D., Liotou, E., Passas, N., Merakos, L.: A survey on parametric QoE estimation for popular services. J. Netw. Comput. Appl. **77**(1), 1–17 (2017)
4. Demirbilek, E., Gregoire, J.-C.: Towards reduced reference parametric models for estimating audio-visual quality in multimedia services. In: IEEE International Conference on Communications (ICC), Kuala Lumpur (2016)
5. Hoßfeld, T., Heegaard, P.E., Varela, M., Moller, S.: QoE beyond the MOS: an in-depth look at QoE via better metrics and their relation to MOS. Qual. User Exp. **1**(2), 1–23 (2016)
6. Yusuf-Asaju, A. W., Zulkhairi D., Ta'a, A.: A proposed analytical customer satisfaction prediction model for mobile internet networks. In: Pacific Asia Conference on Information Systems (PACIS) Proceedings, Langkawi, Malaysia (2017)
7. Xiao, V., Boutaba, R.: Assessing network service profitability: modelling from market science perspective. IEEE/ACM Trans. Netw. **15**(6), 1307–1320 (2007)
8. Djogatovic, V. R., Kostic-Ljubisavljevic, A., Stojanovic, M., Mikavica, B.: Quality of experience in telecommunication. In: 8th International Quality Conference, Kragujevac (2014)
9. Aroussi, S., Mellouk, A.: Statistical evaluation for quality of experience prediction based on quality of service parameters. In: 23rd International Conference on Telecommunications (ICT), Thessaloniki (2016)
10. Battisti, F., Carli, M., Paudyal, P.: QoS to QoE mapping model for wired/wireless video communication. In: Euro Med Telco Conference (EMTC) (2014)
11. Reiter, U., Brunnström, K., De Moor, K., Larabi, M.-C., Pereira, M., Pinheiro, A., You, J., Zgank, A.: Factors Influencing Quality of Experience. Quality of Experience: Advanced Concepts, Applications, pp. 55–72. Springer, Cham (2014)
12. Reichl, P., Egger, S., Möller, S., Kilkki, K., Fiedler, M., Hossfeld, T., Christos, T., Asrese, A.: Towards a comprehensive framework for QOE and user behavior modelling. In: Seventh International Workshop on Quality of Multimedia Experience (QoMEX), Pylos-Nestoras (2015)
13. Tsiaras, C., Stiller, B.: A deterministic QoE formalization of user satisfaction demand (DQX). In: Annual IEEE Conference on Local Computer Networks, Canada (2014)
14. Spiess, J., T'Joens, Y., Dragnea, R., Spencer, P., Philippart, L.: Using big data to improve customer experience and business performance. Bell Labs Tech. J. **18**(4), 3–17 (2014). https://doi.org/10.1002/bltj.21642
15. Ascom: Ascom Network Testing (2012). http://www.livingston-products.com/products/pdf/139777_1_en.pdf
16. Ericson: Basic Concepts of HSPA, 02 (2007). http://escrig.perso.enseeiht.fr/HSPA-Concepts.pdf
17. Galit, S., Peter, B.C., Inbal, Y., Nitin, P.R., Kenneth, K.C.J.: Data Mining for Business Analytics, Concepts, Techniques and Application in R. Wiley, Hoboken (2018)