engineering proceedings

MDPI

*Proceeding Paper*

# Hybrid K-Mean Clustering and Markov Chain for Mobile Network Accessibility and Retainability Prediction †

**Amel Salem Omer [1,*], Tesfaye Addisie Yemer [2] and Dereje Hailemariam Woldegebreal [1]**

1 School of Electrical and Computer Engineering, Addis Ababa University, Addis Ababa P.O. Box 385, Ethiopia; dereje.hailemariam@aait.edu.et
2 Ethio Telecom, Addis Ababa P.O. Box 1047, Ethiopia; tesfaye.addisie@ethiotelecom.et
* Correspondence: amel.salem@yenepay.com

† Presented at the 8th International Conference on Time Series and Forecasting, Gran Canaria, Spain, 27–30 June 2022.

**Abstract:** To provide reliable services, mobile network operators (MNOs) continuously collect vital mobile network performance data to monitor and analyze the functioning of their radio access networks (RANs). RAN is a critical infrastructure for mobile networks and its performance is measured by key performance indicators (KPIs) such as accessibility, retainability, availability, integrity, and mobility. The standard practice is that network managers utilize KPIs to identify failures or unusual events that can significantly degrade the quality of service delivery and the end-users' experiences. However, taking corrective steps based on monitored performance parameters is a reactive approach that contributes to network and service degradations until corrective actions are taken. With the monitoring and automation of RAN infrastructure performance in mind, this paper presents the Markov chain, a widely used probabilistic modeling approach, as a systematic method for jointly predicting network accessibility and retainability status, two of the crucial RAN performance measures. The novel joint prediction is proposed to have a single operation for both accessibility and retainability. Real-time hourly KPIs data was collected from 1530 cells (base stations) run by an operator's network in Addis Ababa, Ethiopia, for 4 months, from 1 November 2020 to 28 February 2021. The cells are scattered across the capital city, where factors such as land use, settlement patterns, and customers behaviors differ. To capture the spatial variation of the KPIs without escalating the computational complexity much, the dataset is separated into six clusters using the K-mean clustering approach. The Markov chain KPIs status prediction models are formulated on a cluster level. The results reveal that the proposed models can predict the KPIs status with 94.61 percent accuracy. Because the data is already available and can be collected at any time using the operator's network management system (NMS), this is a cost-effective technique to proactively improve mobile network performance.

**Keywords:** accessibility; retainability; Markov chain; K-mean clustering



**Citation:** Omer, A.S.; Yemer, T.A.; Woldegebreal, D.H. Hybrid K-Mean Clustering and Markov Chain for Mobile Network Accessibility and Retainability Prediction. *Eng. Proc.* **2022**, *18*, 9. https://doi.org/10.3390/engproc2022018009

Academic Editors: Ignacio Rojas, Hector Pomares, Olga Valenzuela, Fernando Rojas and Luis Javier Herrera

Published: 21 June 2022



## 1. Introduction

The demand for dependable mobile network services is growing and is projected to continue to grow in the coming years. To meet this rising service demand, mobile network operators (MNOs) are expanding their networks and use a centralized network management system (NMS) to monitor the performance of the radio access network (RAN) and core network, two critical components of mobile network infrastructure. NMS is a network monitoring and control tool, with fault management and performance management its two essential functionalities [1]. Fault management is the need for fault-free operation and has three aspects, namely fault identification, fault isolation, and fault correction. The fault identification is conducted with the help of network alarms, while fault isolation of the network's remaining components from the failure is needed so that the isolated network can continue to function normally. Fault correction requires repairing or replacing failed components. Performance management, on the other hand, includes network monitoring to observe network activities and network control to take mitigations that increase network performance. Some of the network manager's performance concerns include determining the amount of capacity utilization, traffic monitoring, throughput status, and reaction time status, among others [1].

Although the NMS offers critical information through various management subsystems, most operators find it challenging to manage the data collected from the system and take corrective actions in a timely manner. Operators select key performance indicators (KPIs) and monitor them at hourly, daily, weekly, or monthly intervals to discover problems or unusual events that might drastically affect service delivery and end-user experience. KPIs are further grouped to assess network performance, and the widely used performance measures are accessibility, retainability, availability, integrity, and mobility. NMS holds vast amounts of historical network performance data, from which possible trends and patterns can be revealed using cutting-edge data mining techniques.

In mobile networks, Markov chain is used for call admission control [2], quality of experience (QoE) modeling [3], quality of service (QoS) modeling [4], efficient resource utilization [5], prediction of user mobility [6], handover management, and network operation status monitoring [7]. In [8,9], Markov chain is proposed to forecast radio resource controller (RRC) setup success and call setup success rates (CSSR) for the Long-term Evolution (LTE) mobile network. The status or state as per Markov's terminology of a cell (base stations) is classified as "Good/High," "Moderate/Acceptable," or "Bad/Low" based on the RRC success rate. Data collected from an operator's network is used to create the Markov chain–based models for RRC and CSSR future state predictions. A given cell is in one of the three states depending on the time of a day, the cell's geographic location, network capacity, and other user- and network-related factors. In addition to the Markov chain, cluster-based approaches, decision trees, and artificial neural networks were employed in [10–12] to estimate a network accessibility-related parameters. These and papers such as [13–15] addressed related performance measures for various generations of cellular mobile systems.

This paper's primary goal is to forecast mobile network accessibility and retainability status using real-time data gathered from the NMS of a major network operator in the capital city of Addis Ababa, Ethiopia. Specifically, the data were collected on an hourly basis from 1530 cells for 4 months' duration, from 1 November 2020 to 28 February 2021. The states of these two critical RAN performance parameters are defined based on the International Telecommunication Union's (ITU's) recommendations for network accessibility and retainability. As the cells are scattered across different geographic regions of the capital city, K-mean clustering technique is used to group cells having spatially correlated performances. The per-cluster averaged data are used to construct the Markov chain prediction model. Two approaches are used for the model formulation, and one is a separate approach so that two Markov models are built for accessibility and retainability. In the joint modeling, a single model is used to predict both parameters. Using either of the two approaches, we can compute the state of the network and the number of transitions until a steady-state is reached. The essential contributions of the research are mentioned here.

- In contrast to prior attempts, we established four states [16], namely "Idle," "Good," "Acceptable," and "Bad" states, to conform to the ITU's recommendations. Furthermore, the Markov chain is constructed to jointly estimate accessibility and retainability in a single operation, yielding a model with 16 states. Four-state separate estimation is employed as a benchmark for comparison. Incorporating ITU's recommendations for state definition and the joint prediction proposal are the unique contributions of this research.
- Previous models only operate for a single cell, leaving out the correlated nature of accessibility and retainability in the spatial domain. Including more cells, however, increases the number of combined states; thus, the Markov model may not scale as the number of cells increases. As an alternative to replicating the prediction method as

many times as the number of cells, we employed K-mean clustering to identify related cells. The Markov chain is then applied to the per-cluster averaged data. Prediction aids in analyzing the status of the considered mobile network.

The remaining paper is organized as follows. Section two discusses fundamental concepts and formulas in accessibility and retainability. Section three introduces some basic concepts of discrete Markov chains. Section four presents and discusses the results obtained. Finally, Section five concludes the paper by identifying possible future directions.

## 2. Accessibility and Retainability KPIs

KPIs obtained from network counters can be grouped into accessibility, retainability, integrity, mobility, and other factors in order to manage and track the performance of the network [17]. According to ITU, service accessibility is "*the ability of a service to be obtained, within specified tolerances and other given conditions, when requested by the user*." Service retainability is "*the ability of a service, once obtained, to continue to be provided under given conditions for a requested duration*" [18].

### *2.1. Accessibility*

The accessibility KPI is expressed in probabilities, which indicate how likely a user is able to access the mobile service during specific service times and conditions. Accessibility measures the network's performance during call setup or before establishing a bearer [19]. For data availability reason, this paper focuses on the Third Generation (3G) mobile networks. RRC, radio access bearer (RAB), Enhanced Universal Terrestrial RAN RAB (ERAB), and CSSR are critical accessibility parameters, as presented below.

- RRC setup success rate (RRC SSR) evaluates the call success rate in a cell or cluster. The formula for this KPI is:

$$RRC\ SSR = \frac{Number\ of\ RRC\ setup\ success}{number\ of\ RRC\ connection\ attempt} \times 100\%. \tag{1}$$

- RAB setup success rate (RAB SSR) evaluates the success rate of assigning a RAB during a call setup procedure. The formula for this KPI is given as follows:

$$RAB\ SSR = \frac{Number\ of\ RAB\ setup\ success}{number\ of\ RAB\ setup\ attempt} \times 100\%. \tag{2}$$

- CSSR is used to evaluate the call setup success at the cell or cluster level. This KPI is calculated based on RRC SSR and RAB SSR for the case of third generation (3G) networks and ERAB SSR for the case of LTE networks.

$$Accessibility = CSSR = RRC\ SSR \times RAB\ SSR \times 100\%. \tag{3}$$

### *2.2. Retainability*

Retainability assesses a network's performance after RAB is established and indicates the proportion of calls that serve the essential service without call drops.

$$Retainability = \left(1 - \frac{Number\ of\ RAB\ abnormal\ release}{d\ total\ number\ of\ RAB\ release}\right) \times 100\%. \tag{4}$$

Equation (4) fraction shows the call drop rate (CDR) value.

## 3. Discrete-Time Markov Chain

A Markov chain is a particular class of a stochastic process with random variables designating the states or outputs of the system [7,20]. The probability of the system transitioning from its current state to a future state depends only on the current state. The collection of states forms a state space of alphabet size $N$. Let $\{a_1, a_2, \dots, a_N\}$ designate the state space and let a sequence of states $S_1, \ldots, S_n \ldots,$ generated by the system in time, where $S_n \in \{a_1,\ a_2,\ \ldots,\ a_N\}$ and $n$ in $S_n$ indicates the discrete-time index.

For the Markov chain fulfilling the memoryless assumption, the transition probability is expressed as [21]:

$$\begin{aligned} P(S_{n+1} &= a_j \middle| S_n = a_i, S_{n-1} = a_h,\ \ S_{n-2} = a_g,\ \ldots.) \\ &= \mathrm{P}\left(S_{n+1} = a_j \middle| S_n = a_i\right), \end{aligned} \tag{5}$$

where $1 \leq i, j \leq N$. We learn from the Markov property that only the most recent state matters to predict the next or future state. From Equation (5), the transition probability from state $a_i$ to state $a_j$ is designated as:

$$P_{ij} = \mathrm{P}\left(S_{n+1} = a_j \middle| S_n = a_i\right). \tag{6}$$

For all $i$ and $j$, the summation of all transition probabilities in a row must be equal to one, i.e.,

$$\sum_{j=1}^{n} p_{ij} = 1. \tag{7}$$

### 3.1. Transition Probability Matrix

The collection of the transition probabilities $P_{ij}$ forms the probability transition matrix (TPM), $P$ (See Equation (8)). Each entry of the matrix shows the probability that the system will transition or remain in the same state. $P$ is a square matrix with the same dimension as the number of states.

$$P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & .. & P_{1N} \\ P_{21} & P_{22} & P_{23} & .. & P_{2N} \\ . & . & . & .. & . \\ . & . & . & .. & . \\ P_{N1} & P_{N2} & P_{N3} & .. & P_{NN} \end{bmatrix} \tag{8}$$

The transition probability $P_{ij}$ is computed from empirical data by counting the number of transitions from state $i$ to state $j$ and dividing the result by the count of all transitions from state $i$ [7].

### 3.2. Initial (Probability or State) Distribution

The initial state distribution is usually expressed as a probability distribution vector, $U$ of dimension $1 \times n$, as shown in Equation (9), with entries that indicate the probability that the system is in a given state at a given initial time. Each entry of the vector is non-negative and the sum of the all entries should be unity.

$$U = [P_1,\ P_2,\ \ldots., P_N]. \tag{9}$$

Without accurate knowledge of the initial distribution, the system can be considered to be in one state with absolute certainty, i.e., probability of unity.

### 3.3. Steady-State Distribution

One of the fascinating aspects of systems that obey the Markov chain is that, after a sufficient number of iterations/transitions, the chain converges to a steady-state, stable, equilibrium, or static distribution [7]. A steady-state condition is one in which the probability of the next state is the same regardless of the present state.

With knowledge of the transition matrix $P$ and the initial probability vector $U$, the probability distribution of the chain after $k$ transitions in the future is given by [7].

$$U^{(k)} = UP^k \tag{10}$$

$P^k$ is the result of multiplying the transition matrix $k$ times by itself. Each element of $U^{(k)}$, designated as $P_{ij}^{(k)}$, is the probability of going from state $i$ to state $j$ in $k$ iteration. As we keep iterating through state transitions by applying $P^k$, the probability vectors $U^{(k)}$ converge to some fixed value, say $\pi_{(k)}$. That is called the *steady-state distribution* and mathematically written in the form as in Equation (11) below.

$$\lim_{k\to\infty} U^{(k)} = UP^k = \pi_{(k)}. \tag{11}$$

We note from Equation (11) that the Markov chain probabilistically predicts the system's future state based on knowledge of state space, initial distribution, and transition matrix.

### 3.4. Transition Diagram

A transition diagram, which illustrates all of the system's transitions, is another way to display the TPM. A directed arrow shows the presence of a transition from one state to another state, and each node represents a state of the Markov chain. The edge represents the current state, and the arrow points towards the next state [7].

## 4. Results and Discussions

This section covers data collection and accessibility and retainability status prediction using a four- and sixteen-state Markov chain model.

### 4.1. Data Collection and Preprocessing

The performance report system (PRS) installed in the operator's network was used to collect real-time hourly data from 1530 cells for 4 months' duration.

- Linear interpolation is used to fill data gaps caused by factors such as cell outages and connection problems among cells and central radio network controller (RNC).
- If no voice or data service attempts are made in a cell for one hour, the accessibility and retainability values are zero. This situation is handled separately, and the accessibility or retainability status is "Idle." Figure 1 shows RRC and RAB attempt values for Cluster 6 throughout a week, and both values are zero at midnight.

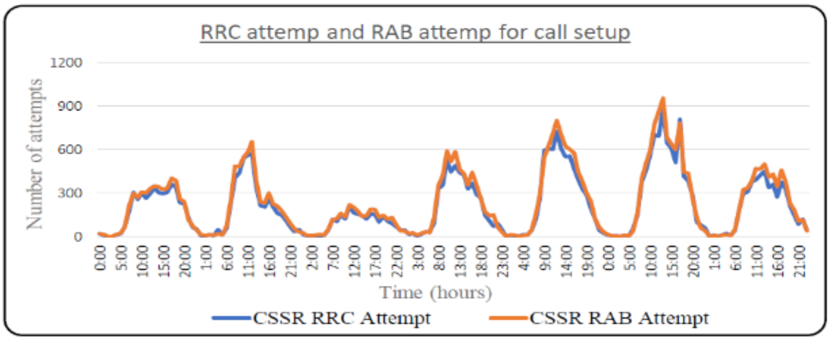


**Figure 1.** One-week RRC and RAB attempts.

- The data are split into two, with 60% utilized for training and 40% for model validation/testing. The training data are used to generate the transition matrix, and the process of constructing such a matrix from data are described in [7]. Combinations of 70/30 and 80/20 are also utilized for comparison purposes.
- The system predicts the next probability vector given the current state probability distribution and the transition matrix. The operation is then continued until a steady-state condition is reached. Following step/iteration prediction, results are compared to the validation data to assess prediction accuracy.

### 4.2. Clustering

It takes time to analyze individual cell performance and patterns. In this research, we suggest K-mean clustering as a method for grouping cells with similar accessibility and retainability properties. Model construction and prediction are based on per-cluster averaged accessibility and retainability. The Elbow approach in K-mean clustering is used to identify the number of clusters by changing the parameter from 2 to 18. Each cell was randomly assigned to several clusters to vary the centroid of each center. The procedure was repeated until the cluster variation in the data could no longer be reduced by adjusting the cluster centroid. We discovered that a clustering value of 6 is adequate. Hourly data acquired from each cell varies from 0% to 100%; however, if no voice or data service requests are received in a cell for 1 h, all counter values for that hour are zero, as illustrated in Figure 1.

### 4.3. KPI Threshold for States Definition

Operators set threshold values for several KPIs based on the ITU's recommendations, considering variables such as capital expenditures, operational expenses, QoS, and customer satisfaction. Tables 1 and 2 display a threshold value for the considered operator's accessibility and retainability. Based on the values in the two tables, the states of accessibility and retainability are generated.

**Table 1.** Possible values of call setup attempt and CSSR.

| Call Setup Attempt | Value | State of a Cell |
|---|---|---|
| >0.0 | CSSR $\geq$ 98.0% | Good (G) |
| >0.0 | 95.0% $\leq$ CSSR $\leq$ 98.0% | Acceptable (A) |
| >0.0 | 0.0% $\leq$ CSSR $\leq$ 95.0% | Bad (B) |
| =0.0 | - | Idle (I) |

**Table 2.** Possible values of RAB setup success and CDR.

| RAB Setup Attempt | Value | State of a Cell |
|---|---|---|
| >0.0 | 0.0% $\leq$ CDR $\leq$ 1.0% | Good (G) |
| >0.0 | 1.0% $\leq$ CDR $\leq$ 3.0% | Acceptable (A) |
| >0.0 | 3.0% $\leq$ CDR | Bad (B) |
| =0.0 | - | Idle (I) |

### 4.4. Separate Prediction

As indicated above, the accessibility and retainability predictions at the cluster level can be made separately and jointly. Four states are required for the separate case. Hence, the corresponding transition matrices are $4 \times 4$. The state transition probability diagram for the sixth cluster is given in Figure 2 below, which is obtained after developing the model.

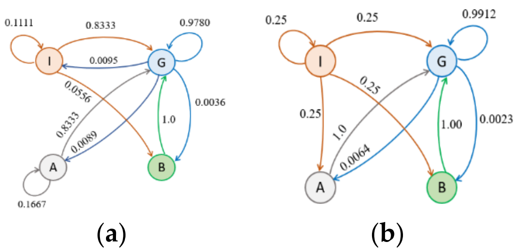


**Figure 2.** Transition probability diagram of cluster 6. (**a**) Accessibility. (**b**) Retainability.

Note that there are missing arrows in the two figures. As an example, there is no arrow in Figure 2a pointing from state A to state I, indicating that such a transition does not exist or the system has never landed in an idle state if it was initially in the Acceptable state.

### *4.5. Joint Prediction*

The different state combinations of accessibility and retainability can be seen via joint estimation. For example, a Bad state of in accessibility and a Bad state in retainability can occur at the same time. When all possible combinations are considered, the total number of states rises to 16, and the resulting transition matrix is shown in Figure 3.

$$
\begin{array}{cc}
 & \begin{array}{cccccccccccccccc} \mathrm{II} & \mathrm{IG} & \mathrm{IA} & \mathrm{IB} & \mathrm{GI} & \mathrm{GG} & \mathrm{GA} & \mathrm{GB} & \mathrm{AI} & \mathrm{AG} & \mathrm{AA} & \mathrm{AB} & \mathrm{BI} & \mathrm{BG} & \mathrm{BA} & \mathrm{BB} \end{array} \\
\begin{array}{c} \mathrm{II} \\ \mathrm{IG} \\ \mathrm{IA} \\ \mathrm{IB} \\ \mathrm{GI} \\ \mathrm{GG} \\ \mathrm{GA} \\ \mathrm{GB} \\ \mathrm{AI} \\ \mathrm{AG} \\ \mathrm{AA} \\ \mathrm{AB} \\ \mathrm{BI} \\ \mathrm{BG} \\ \mathrm{BA} \\ \mathrm{BB} \end{array} &
\left(\begin{array}{cccccccccccccccc}
0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 \\
0.0000 & 0.1176 & 0.0000 & 0.0000 & 0.0000 & 0.7647 & 0.0588 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0588 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 \\
0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 \\
0.0000 & 0.0090 & 0.0006 & 0.0000 & 0.0000 & 0.9707 & 0.0048 & 0.0024 & 0.0000 & 0.0090 & 0.0000 & 0.0000 & 0.0000 & 0.0030 & 0.0006 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.8333 & 0.0000 & 0.0000 & 0.0000 & 0.1667 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 \\
0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 \\
0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 1.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\
0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625 & 0.0625
\end{array}\right)
\end{array}
$$

**Figure 3.** Sixteen-state TPM of cluster 6.

### *4.6. State Prediction*

After creating the transition matrices and knowing the current/initial state distribution, the next state and steady state distributions are predicted using Equation (10). If the current state is (assumed to be) in a Good state, then the value of $\pi_0$ is,

$$\pi_0 = [\ I\ G\ A\ B\ ] = [\ 0\ 1\ 0\ 0\ ] \tag{12}$$

Then, using Equations (10) and (12), the next accessibility probability is $\pi_1$ for one of the clusters when computed, and the result is:

$$
\begin{aligned}
[0\ 1\ 0\ 0] \quad \times & \begin{pmatrix} 0.1111 & 0.8333 & 0.0000 & 0.0556 \\ 0.0095 & 0.9780 & 0.0089 & 0.0036 \\ 0.0000 & 0.8333 & 0.1667 & 0.0000 \\ 0.0000 & 1.0000 & 0.0000 & 0.0000 \end{pmatrix} \\
= & \left[\ 0.0095 \quad 0.9780 \quad 0.0089 \quad 0.0036\ \right]
\end{aligned}
\tag{13}
$$

According to the result, the system has a 0.95 percent chance of going to the Idle state, a 97.80 percent chance of staying in a Good state, a 0.89 percent chance of going to the Acceptable state, and a 0.36 percent chance of going to the Bad state.

Equation (11) is used to find the steady-state distribution calculated iteratively until the next and previous state values are equal. Tables 3 and 4 display the steady-state results for the four-state Markov chain regarding accessibility and retainability. Table 5 depicts the cluster 1 steady-state distribution using the sixteen-state Markov chain. For both scenarios, 70% of the data are used as a training set.

**Table 3.** Steady-state vector of accessibility using four-state Markov chain.

| Cluster | Steady-State Vector of Accessibility | | | |
|---|---|---|---|---|
| | I | G | A | B |
| 1 | 0.0000 | 0.9926 | 0.0060 | 0.0015 |
| 2 | 0.0000 | 0.9722 | 0.0149 | 0.0129 |
| 3 | 0.0000 | 0.9911 | 0.0079 | 0.0010 |
| 4 | 0.0000 | 0.9782 | 0.0179 | 0.0040 |
| 5 | 0.0000 | 0.9916 | 0.0055 | 0.0030 |
| 6 | 0.0109 | 0.9722 | 0.0114 | 0.0055 |

**Table 4.** Steady-state vector of retainability using four-state Markov chain.

| Cluster | Steady-State Vector of Retainability | | | |
|---|---|---|---|---|
| | I | G | A | B |
| 1 | 0.0000 | 0.9980 | 0.0020 | 0.0000 |
| 2 | 0.0000 | 0.9955 | 0.0035 | 0.0010 |
| 3 | 0.0000 | 1.0000 | 0.0000 | 0.0000 |
| 4 | 0.0000 | 0.9980 | 0.0020 | 0.0000 |
| 5 | 0.0000 | 0.9965 | 0.0030 | 0.0005 |
| 6 | 0.0000 | 0.9901 | 0.0079 | 0.0020 |

**Table 5.** Steady-state vector of Cluster 1 using sixteen-state Markov chain.

| Cluster | Steady-State Vector | | | |
|---|---|---|---|---|
| | [II IG IA IB GI GG GA GB AI AG AA AB BI BG BA BB] | | | |
| 1 | [0.0000 | 0.0000 | 0.0000 | 0.0000 |
| | 0.0000 | 0.9926 | 0.0000 | 0.0000 |
| | 0.0000 | 0.0040 | 0.0020 | 0.0000 |
| | 0.0000 | 0.0015 | 0.0000 | 0.0000] |

In Table 3, the maximum value in the Good state from the six clusters is 99.26% in cluster 1, and the minimum value is 97.22% in clusters 2 and 6. The maximum value of the Bad state is 1.29% in cluster 2, and the minimum value is 0.1% in cluster 3. From this cluster, one cell is at the top in the Good state, and cluster 2 cells are at the top in the Bad state. Though cluster 6 cells are the least in the Good state, they are not at the top in the Bad state because, next to the Good state, cluster 6 cells have a high probability (1.09%) of being in the Idle state. So, if optimization or maintenance work is needed, the schedule and priority should be given based on the steady-state vector values of each cluster.

Steady state distribution for the sixteen-state Markov chain follows the same approach. Cluster 1's steady-state outcome is shown in Table 5. The first letter stands for accessibility, while the second stands for retainability, and 99.26% of the time, accessibility and retainability were in the Good state, while for 0.4% of the time accessibility was in the Acceptable state and retainability was in the Good state. Furthermore, for 0.2% of the time, accessibility and retainability were both in the Acceptable state, while 0.15% of the time, accessibility was Bad, and retainability was in the Good state. As a result, the table provides cell information relating to accessibility and retainability, allowing operators to quickly sort cells that perform poorly in either or a combination of the two performance measures.

### 4.7. Evaluation Metric

The accuracy of a model was assessed using Equation (14) [21], which calculates the percentage of correctly forecasting the next state given the current state.

$$Accuracy = \frac{Correct\ predictions}{Total\ number\ of\ examples} \times 100\%. \quad (14)$$

Table 6 below shows the accuracy results for different combinations of training data proportion, clusters, four vs sixteen states modeling and the two KPIS considered. As an example, we note that a minimum value of 96.09% prediction accuracy is achieved in cluster 2 in predicting accessibility when 60% training set is used, while 96.87% prediction accuracy is achieved in cluster 5 in predicting retainability when the 80% training set is used. A 94.61% prediction accuracy is achieved in cluster 6 in predicting both accessibility and retainability when 80% of the data are used for training and when the modeling is the case of the sixteen-state Markov chain.

**Table 6.** Prediction accuracy for sixteen-state Markov chain.

| Cluster | Training Set | Accessibility Accuracy Using Four States | Retainability Accuracy Using Four States | [(Column 3 × Column 4)/100] (%) | Accessibility and Retainability Accuracy Using Sixteen States |
|---|---|---|---|---|---|
| 1 | 60% | 98.7837 | 99.1312 | 97.9254 | 97.8280 |
| | 70% | 98.3796 | 98.8426 | 97.2410 | 97.1065 |
| | 80% | 98.0870 | 98.2609 | 96.3811 | 96.1739 |
| 2 | 60% | 96.0904 | 98.6968 | 94.8381 | 95.5691 |
| | 70% | 96.7593 | 98.3796 | 95.1914 | 96.1806 |
| | 80% | 96.1739 | 97.7391 | 93.9995 | 95.4783 |
| 3 | 60% | 98.5230 | 100.0000 | 98.5230 | 98.5230 |
| | 70% | 98.4954 | 100.0000 | 98.4954 | 98.4954 |
| | 80% | 98.6087 | 100.0000 | 98.6087 | 98.6087 |
| 4 | 60% | 98.5230 | 100.0000 | 98.5230 | 98.5230 |
| | 70% | 98.3796 | 100.0000 | 98.3796 | 98.3796 |
| | 80% | 98.6087 | 100.0000 | 98.6087 | 98.6087 |
| 5 | 60% | 97.7411 | 98.4361 | 96.2126 | 97.0460 |
| | 70% | 97.3380 | 97.9167 | 95.3101 | 96.4120 |
| | 80% | 96.1739 | 96.8696 | 93.1633 | 94.7826 |
| 6 | 60% | 96.5248 | 98.0886 | 94.6798 | 94.7871 |
| | 70% | 96.8750 | 98.0324 | 94.9689 | 95.0231 |
| | 80% | 96.6957 | 97.7391 | 94.5095 | 94.6087 |

## 5. Conclusions

In this paper, the two important mobile network KPI parameters of accessibility and retainability are predicted by formulating the Markov chain in four states and sixteen states. The sixteen-state Markov chain is formulated in a bid to jointly estimate both KPIs in a single operation. Moreover, in order to capture the spatial behaviors of these KPIs, K-mean clustering is applied to cluster the data from 1530 cells into 6 clusters. States are created based on threshold values set by operators and the developed models are validated by splitting the data for training and testing. We hope the approach provides significant insight on how to use data available within an operator's NMS to better understand the status of a network.

This work might be improved in some ways. Conducting the prediction for a large number of cells in a computationally efficient manner and to obtain per-cell level information is one research area. The clustering and joint approach may not scale well as the number of cells grows. Moreover, applying the approach for other KPIs, network types, and services is an area worth exploring. Finally, future research should employ the hidden Markov model for status modeling and prediction.

**Author Contributions:** Conceptualization, T.A.Y. and D.H.W.; methodology, T.A.Y. and D.H.W.; software, T.A.Y. and A.S.O.; validation, A.S.O. and T.A.Y.; formal analysis, A.S.O., T.A.Y. and D.H.W.; investigation, A.S.O., T.A.Y. and D.H.W.; resources, T.A.Y.; data curation, T.A.Y.; writing—original draft preparation, A.S.O. and T.A.Y.; writing—review and editing, A.S.O. and D.H.W.; visualization, A.S.O.; supervision, D.H.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Restrictions apply to the availability of the data used for the research. Data was obtained from ethio telecom and are available from the second author with the permission of ethio telecom.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Yemer:, T.A. Mobile Networks Accessibility, and the Retainability States Prediction Using Markov Chain. Master's Thesis, Addis Ababa University, Addis Ababa, Ethiopia, January 2022.
2. Huawei Technologies Co., Ltd. *HUAWEI RAN KPI for Performance Management (RNC V100R006)*; Huawei Technologies Co., Ltd.: Shenzhen, China, 2006.
3. Machine Learning Glossary. Available online: https://developers.google.com/machine-learning/glossary#multi-class (accessed on 19 May 2021).
4. International Telecommunication Union. *Quality of Service and Dependability Vocabulary*; The International Telegraph and Telephone Consultative Committee: Geneva, Switzerland, 1988; Volume E.800, p. 16.
5. Tolver, A. *An Introduction to Markov Chain*; Department of Mathematical Sciences, University of Copenhagen: Copenhagen, Denmark, November 2016.
6. Ng, H.Y.; Ko, K.T.; Tsang, K.F. 3G Mobile Network Call Admission Control Scheme Using Markov Chain. In Proceedings of the Ninth International Symposium on Consumer Electronics, Macau, China, 14–16 June 2005. [CrossRef]
7. Mitra, K.; Ahlund, C.; Zaslavsky, A. QoE Estimation and Prediction Using Hidden Markov Models in Heterogeneous Access Networks. In Proceedings of the Australasian Telecommunication Networks and Applications Conference (ATNAC), Brisbane, Australia, 7–9 November 2012. [CrossRef]
8. Wu, Q.; Zhang, M.; Zheng, R.; Lou, Y.; Wei, W. A QoS-satisfied Prediction Model for Cloud-Service Composition Based on a Hidden Markov Model, Special issue on Applied Mathematics and Algorithms for Cloud Computing and IoT, Hindawi, July 2013. *Math. Probl. Eng.* **2013**, *2013*, 387083. [CrossRef]
9. Hendrawan. RRC success rate accessibility prediction on SAE/LTE network using Markov chain model. In Proceedings of the 2017 11th International Conference on Telecommunication Systems Services and Applications (TSSA), Lombok, Indonesia, 26–27 October 2017. [CrossRef]
10. Amirrudin, N.A.; Ariffin, S.H.S.; Abd Malik, N.N.N.; Ghazali, N.E. Mobility Prediction via Markov Model in LTE Femtocell. *Int. J. Comput. Appl.* **2013**, *65*, 18.
11. Wang, Y. Evaluating Wireless Network Accessibility Performance Via Clustering-Based Model: An analytic methodology. In Proceedings of the Wireless Telecommunications Symposium (WTS), London, UK, 18–20 April 2016. [CrossRef]
12. Hendrawan, N.A. Accessibility Degradation Prediction on LTE/SAE Network Using Discrete-Time Markov Chain (DTMC) Model. *J. ICT Res. Appl.* **2019**, *13*, 1–18. [CrossRef]
13. Iyer, A.P.; Li, L.E.; Stoica, I. Automating Diagnosis of Cellular Radio Access Network Problems. In Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking, Snowbird, UT, USA, 16–20 October 2017. [CrossRef]
14. Oluwafemi, E.O.; Nnonye, O.E.; Okechukwu, U.; Adewale, A.L. Prediction of Call Drops in GSM Network using Artificial Neural Network. *J. Teknol. Syst. Comput.* **2019**, *7*, 38–46.
15. Akanbasiam, J.A.; Ngala, D.K. The Study of Quality of Service on a Major Mobile Network Operator in Ghana. *IOSR J. Electron. Commun. Eng. (IOSR-JECE)* **2017**, *12*, 21–25. [CrossRef]

16. Olayinka, O.O.; Olukemi, S.O.; Chukwuemeka, O. Assessment of Quality of Service of Mobile Network Operators in Akure. *Int. J. Bus. Adm.* **2019**, *10*, 3. [CrossRef]
17. ETSI TS 132 410, Version 9.0.0, Release 9. 2010. Available online: https://cdn.standards.iteh.ai/samples/32882/33cf7e8727da4bf7bb448687b19604f9/ETSI-TS-132-410-V9-0-0-2010-01-.pdf (accessed on 19 May 2022).
18. Porod, U. Dynamics of Markov Chains for Undergraduates. 19 February 2021. Available online: https://www.math.northwestern.edu/documents/book-markov-chains.pdf (accessed on 17 June 2022).
19. Abdulkareem, H.A.; Tekanyi, M.S.; Kassim, A.Y.; Zakariyya, Z.M.; Almustapha, M.D.; Abdu-Aguye, U.F.; Adamu, H. Analysis of a GSM Network Quality of Service Using Call Drop Rate and Call Setup Success Rate as Performance Indicators. *Eur. J. Electr. Eng.* **2020**, *9*, 113–121.
20. Omer, A.S.; Woldegebreal, D.H. Review of Markov Chain and Its Applications in Telecommunication Systems. In *E-Infrastructure and E-Services for Developing Countries*; Sheikh, Y.H., Rai, I.A., Bakar, A.D., Eds.; AFRICOMM 2021; Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering; Springer: Cham, Switzerland, 2022; Volume 443. [CrossRef]
21. Paranthaman, V.V.; Mapp, G.; Shah, P.; Nguyen, H.X.; Ghosh, A. Exploring Markov Models for the Allocation of Resources for Proactive Handover in a Mobile Environment. In Proceedings of the IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops), Clearwater Beach, FL, USA, 26–29 October 2015. [CrossRef]