

Article

Not peer-reviewed version

---

# Trends and Challenges in Data Analytics and Machine Learning Using Call Detail Records in Telecom Systems

---

[James Barrett](#)<sup>\*</sup>, [Phil Legg](#), [Jim Smith](#), Tim Barnes, Charles Boyle

Posted Date: 17 September 2024

doi: 10.20944/preprints202409.1334.v1

Keywords: call detail records; big data; mobile networks; machine learning



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## Article

# Trends and Challenges in Data Analytics and Machine Learning Using Call Detail Records in Telecom Systems

James Barrett <sup>1,\*</sup>, Phil Legg <sup>1</sup>, Jim Smith <sup>1</sup> Tim Barnes <sup>2</sup> and Charles Boyle <sup>2</sup>

<sup>1</sup> University of the West of England, Bristol, UK

<sup>2</sup> Ribbon Communications Ltd.

\* Correspondence: james6.barrett@uwe.ac.uk

**Abstract:** As telecommunication networks continue to grow, the volume, variety, and velocity of such 'big data' poses an immense challenge for security and service analysts. A primary data format in such networks are Call Detail Records (CDRs), and more recently, eXtended Detail Records (XDRs), that capture a variety of information such as caller, receiver, duration, carrier, datetime, amongst other attributes. In this short paper, we investigate the current research trends and challenges that exist within this data-rich domain. As part of this, we look at the domain-specific tasks that analysts will seek to address, we look at how machine learning has been utilised to date within this domain, and we present our current findings on open research problems that are ripe for further investigation.

**Keywords:** call detail records; big data; mobile networks; machine learning

## 1. Introduction

Modern telecommunications systems are highly data-rich platforms that include mobile phone usage and Internet services access that have shaped our global society. At the core of such services, telecommunications systems gather Call Detail Records (CDR) and eXtended Detail Records (XDR) for capturing detailed logs related to the communications taking place. Given the sheer scale of data being generated globally every minute, analysis of such data is a challenging and laborious task. Furthermore, being able to conduct this analysis in real-time to support aspects of security and service are vital. In this short paper, we will explore the current trends and challenges within modern telecommunications systems. In particular, we focus on how data analytics and machine learning techniques have been applied to applications that investigate CDR and XDR data, ranging from the analysis of call behaviours through to the contextualisation of further activities based on what the underlining CDR and XDR information may reveal about users. We summarise our findings through the identification of open research challenges that may provide scope and direction for other researchers.

## 2. Background

### 2.1. Definitions of CDR & XDR

Traditional data records about telecommunications transactions are captured through the use of Call Detail Records (CDR). As one may imagine, these records will document attributes such as time, duration, status, source user (e.g., phone number) and destination number to summarise the activity. For mobile telecommunications, CDR may also refer to the Charging Data Record, as recognised by the European Telecommunications Standards Institute (ETSI) and the European Standards Organisation (ESO). ETSI TS 132 298<sup>1</sup> provides the full standard of the Charging Data Record parameter description for use within 3GPP (3rd Generation Partnership Project (3GPP)) mobile telecommunication networks. Here, the primary usage of CDR is for charging of service and so interoperability across networks is

<sup>1</sup> [https://www.etsi.org/deliver/etsi\\_ts/132200\\_132299/132298/15.10.00\\_60/ts\\_132298v151000p.pdf](https://www.etsi.org/deliver/etsi_ts/132200_132299/132298/15.10.00_60/ts_132298v151000p.pdf)

key, compared to eXtended Data Records (XDR) which are vendor or application-specific. Entities in the CDR are captured as name-value pairs, and may quantify usage based on bytes of data download or duration of data download, as well as information about which base station has logged the call transaction. As the name suggests, XDR provides an extension on the CDR concept, allow for greater detail and customisation in terms of the specific parameters that are captured by a facility or logging device, meaning that as usage of telecommunications systems have evolved, for example usage of mobile Internet services, systems have been able to adapt to account for new forms of data logging whilst offering similarity in data formats with legacy systems.

**Table 1.** Example Of CDR Format

Header Field Name	Value
Caller ID	101032
Caller IMSI	12039810283012
Caller Carrier	Verizon
Receiver ID	101303
Receiver IMSI	04920398402302
Receiver Carrier	Three
Duration (Minutes, Seconds)	30:12
Call Initiated	2-50/18/02/2022
Call Terminated	3-20/18/02/2022

The lightweight nature of CDR/XDR call records is crucial given that a mobile network operator could easily be capturing several million records for as little as a single day. Nevertheless, they provide a wealth of information for analysts in terms of service operability data.

## 2.2. Dataset Availability and Considerations

In order to fully assess the suitability of machine learning and data science concepts for understanding telecommunications activity, obtaining suitable example data is a fundamental requirement for success. In terms of academic research and existing public datasets, the NoDoBo project [1] by the University of Strathclyde has been commonly used by researchers. This dataset provides mobile phone usage of 27 high-school students, from September 2010 to February 2011, and includes data for 13035 call records, 83542 message records, 5292103 presence records, and other related data. The NoDoBo dataset has since been made publicly-available via the CRAWDAD data repository<sup>2</sup>. Researchers including Sultan et al. [2] have used this dataset to examine anomaly detection and traffic prediction tasks, using k-means clustering and ARIMA time-series modelling.

The *churn* data set, originally provided by Blake and Merz<sup>3</sup>, examines 15 variables about user call usage as derived from CDR data such as the number of minutes, number of calls and number of charges for different periods of the day (i.e., Day, Evening, Night), as well as for International usage, in order to predict user churn. This dataset has been used by various researchers, including Brandusoiu and Todorean [3], who have since examined the use of Principal Component Analysis (PCA) for modelling lower dimensional representation of the parameter space. In their investigation, they reduce 15 parameters down to 11, suggesting that actually the majority of parameters are indeed important in order not to lose information within the parameters space.

The Iranian Churn Dataset, available from the UCI Machine Learning repository<sup>4</sup>, is randomly collected from an Iranian telecom company database over a period of 12 months. A total of 3150 rows of data, each representing a customer, bear information for 13 columns. The attributes that are in this dataset are call failures, frequency of SMS, number of complaints, number of distinct calls, subscription

<sup>2</sup> <https://crawdad.org/strath/nodobo/20110323/>

<sup>3</sup> <https://www.kaggle.com/becksddef/churn-in-telecoms-dataset>

<sup>4</sup> <https://archive.ics.uci.edu/ml/datasets/Iranian+Churn+Dataset>

length, age group, the charge amount, type of service, seconds of use, status, frequency of use, and Customer Value. All of the attributes except for attribute churn is the aggregated data of the first 9 months. The churn labels are the state of the customers at the end of 12 months. The three months is the designated planning gap. This dataset has most recently been used in research by Jafari-Marandi et al. [4] that explores the use of self-organizing error-driven artificial neural networks, to combine unsupervised and supervised learning techniques for the purpose of churn prediction.

Whilst above we discuss three real-world telecom datasets used by the research community, the use of real-world datasets naturally pose a number of challenges such as privacy and anonymisation of users, relevancy to the intended research task, and granting access from a mobile operator. Furthermore real-world datasets are time-consuming to curate under controlled research conditions, and can become out-dated, or limited in their general usage, fairly quickly. Songailaite and Krilavicius [5] propose a synthetic CDR data generation to overcome many of these challenges. Their approach uses statistical analysis to extract the users' behaviour from a real-life dataset, therefore being able to scale up the underlying characteristics of users exhibited within the synthetic data. In their research, they identified four key areas that are considered to be problematic in telecoms research: 1) increased load over an entire area, 2) increased load at a specific location, 3) natural disasters, causing increased load and increased call failure, and 4) cell tower failure, similar to that of point (3). For the purpose of their research they simulate scenarios (1) and (4). Figure 1 shows an example output from their CDR generator tool. They provide the CDR generator as open-source to the research community<sup>5</sup>, and they include details of injecting the two scenarios as described above, providing a suitable basis to extend this functionality further.

Caller id	Caller compar	Receiver id	Receiver company	Timestamp	Duration s.
97394	Kompanija1	100599	Kompanija1	04/12/2020 01:52	2038
70684	Kompanija1	49594	Kompanija2	04/12/2020 06:50	8
75168	Kompanija1	82076	Kompanija1	04/12/2020 07:01	137
72845	Kompanija1	52382	Kompanija3	04/12/2020 07:04	1169
57901	Kompanija1	91843	Kompanija1	04/12/2020 07:09	157

Figure 1. Example output from a CDR Generator [5]

### 2.3. Wider Research Using CDR Data

CDR data provides a data-driven view of higher-level social interactions between users. For this reason, CDR data has also been used in research related to social mobility, social networks, and behavioural analysis. Leo et al. [6] study usage and mobility of users in Mexico using CDR data collected 12 months that contains 8 millions users and 5 billions of call events. In particular, they assess user movements between consecutive calls as determined by base station connectivity. Rhoads et al. [7] measure social connections and segregations using CDR, based on real world factors of communications between refugees among the Turkish population. Pinter et al. describe a hybrid machine learning approach for modelling house prices using CDR data [8], whilst Kozik et al. use visualisation techniques for analysing CDR data to predict criminal activity [9]. Whilst these example

<sup>5</sup> <https://github.com/CARD-AI/CDR-Generator>

are not the primary focus of our research, they do illustrate the significant social connection between mobile telecommunications data and the wider societal influences and impact that they can help to model. It is therefore important to recognise that through improved analysis of telecommunications data, these data providers could well have a wider role to play in terms of supporting society.

Abba et al. [10] investigate the issue of forensic analysis on Multiple Mobile Network (MMN) CDR data. A key aspect of this research places value on the integrity of call data records being used as a forensic metric, due to the information being stored as a property of a user and therefore this information is not made available outside of administrative/operator access. They suggest that the MMN-CDR, when forensically analyzed, reveals an increase in time efficiency over the existing CDR due to its high absolute speed, whilst also obtaining higher accuracy and completeness.

Nair et al. [11] conduct an analysis of CDR research for the period 2004-2016. They identify six broad categories: traffic analysis, natural calamities, disease disaster management, call drop, anomaly detection and other applications. It can also be observed from this research that client behaviour classification is used as a metric to provided extended detail in the mapping of user activity for analytical purposes.

We have identified a number of datasets that are utilised by researchers, including both real-world and synthetic cases, and we have discussed some of the challenges associated with each. We have also examined a number of research areas for the analysis of CDR, ranging from a traditional network traffic analysis, anomaly detection, traffic prediction and system maintenance perspective, through to the wider societal implications that can be made based on human interactions using mobile telecommunications networks.

3. Analysis of CDR/XDR

In this section, we delve further into the analysis of existing CDR/XDR datasets to uncover their structure and how machine learning concepts could be applied, investigation both existing analytical approaches and proposing new areas of research. To motivate this section, we examine the NoDoBo dataset further. Figure 2 provides an extract of the NoDoBo dataset as rows, whilst Figure 3 shows a time-series of the call duration data. From the time-series plot we can start to identify anomalous activity in the data, such as when high peaks of activity occur. We can utilise machine learning methods to perform a range of tasks on this data, such as clustering to identify similar groupings of data observations, or we could use regression techniques to perform time-series forecasting and predict how this time-series will continue for future observations. To understand what the true underlying distributions of this data are, further analysis may also be required to assess anomalies in the context of individual users, call recipients, time of day, or some other attributes that may account for the characteristics of the data.

	user	other	direction	duration	timestamp
0	7610039694	07434677419	Incoming	211	Wed Sep 15 19:17:44 +0100 2010
1	7641036117	01666472054	Outgoing	31	Mon Feb 11 07:18:23 +0000 1980
2	7641036117	07371326239	Incoming	45	Mon Feb 11 07:45:42 +0000 1980
3	7641036117	07681546436	Outgoing	10	Mon Feb 11 08:04:42 +0000 1980
4	7641036117	07681546436	Outgoing	0	Mon Feb 11 08:05:31 +0000 1980

Figure 2. Extract from the calls table in the NoDoBO dataset (Not Sorted)



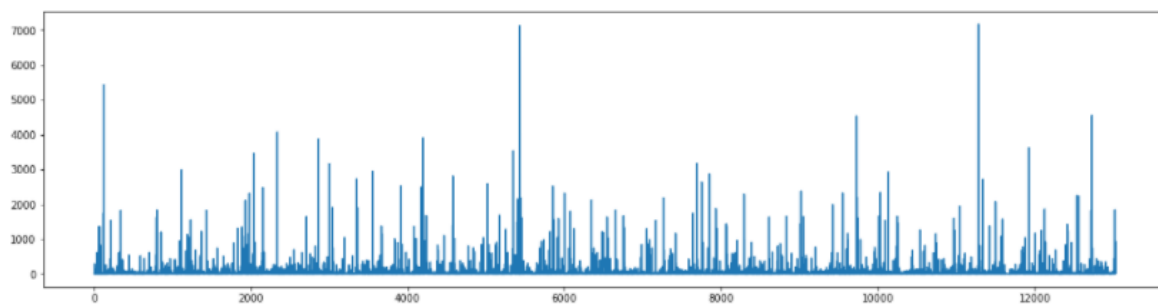


Figure 3. Time-series plot of call duration data in the NoDoBO dataset

The following sections consider different challenges associated with machine learning in the context of CDR data.

**Machine To Machine (M2M)** Whilst data analysis is often completed in a number of different ways, previous research has attempted to use M2M (Machine to Machine) learning techniques [12] to complete prediction using classification tasks with a prediction accuracy of 77% in real application tasks. M2M has been chosen in this research example to be an effective method of multi-layer (application layer to perception-layer) analysis due to end-to-end encryption applications causing issues to real-time analysis. With a high accuracy rate in prediction tests, this approach could detect anomalous call data records with a higher accuracy than other ML methods.

**Hybrid Approaches** Research conducted by [13] into anomaly detection using call data records proposes using a number of different means to analyse CDR information. In contrast to research conducted by [2], this paper attempts to apply a hybrid approach in detection means by applying a series of “anomaly detection methods such as GARCH, k-means, and Neural Networks” [13]. Using methods such as neural networks for the task of prediction and classification could potentially result in higher accuracy rates for anomaly detection and improved user clusters within call data records.

Usage of a hybrid-approach may potentially offer more accurate results in anomaly detection models whereby taking a singular means (as seen in [2] for example) may only allow for a singular means of classification and therefore may potentially miss key insights to the call data records.

**Robust Fuzzy Clustering** Using pseudo-anonymised fixed data [14] Kaur and Ojha propose using a robust fuzzy clustering means to analyse CDR data. This research uses “Fuzzy Logics and LD-ABCD” [14] for knowledge discovery purposes of CDR information, by creating features of previously unknown data. This paper also makes reference to CDR information being a form of spatial dynamics of human communities, as this usage has been noted above in [14] as a common usage of CDR analysis. Applying robust fuzzy clustering to call data record groups may give a potentially higher accuracy rate in classifying specific user groups compared to other approaches such as k-means clustering.

**Temporal and Spatial Analysis** In a paper addressing base station behaviour, data analysis methods for CDR analysis including temporal and spatial analysis [15] are two given techniques for conducting CDR analysis, to produce insights into mobile base station behaviour. It is discussed that temporal and spatial analysis is used to find patterns of data that are considered to have dimensionality. It is also determined through this paper that such insights as primary usage and calling patterns can be determined through temporal and spatial analysis techniques [15]. Similar as to how M2M and k-means can be used for classification, temporal and spatial analysis could also be applied to anomaly detection or prediction of call records, and compared to determine which ML method is most effective.

**Spatiotemporal Resolution & Additional Methods** Khaefi et al. [16] model wealth from call data records & survey data, based on predictive analytics using spatiotemporal resolution methods. It is observed in this paper, that hyperparameter tuning is used of such methods as SVM (Support vector machines), Naive Bayes, Elastic-Net, Neural Networks & Decision Trees to provide the most suitable model [16].

**Semi-Supervised Learning** Jaffry et al. propose a semi-supervised algorithm that is used for the analysis of call data records [17]. This paper provides the context that future self organising networks

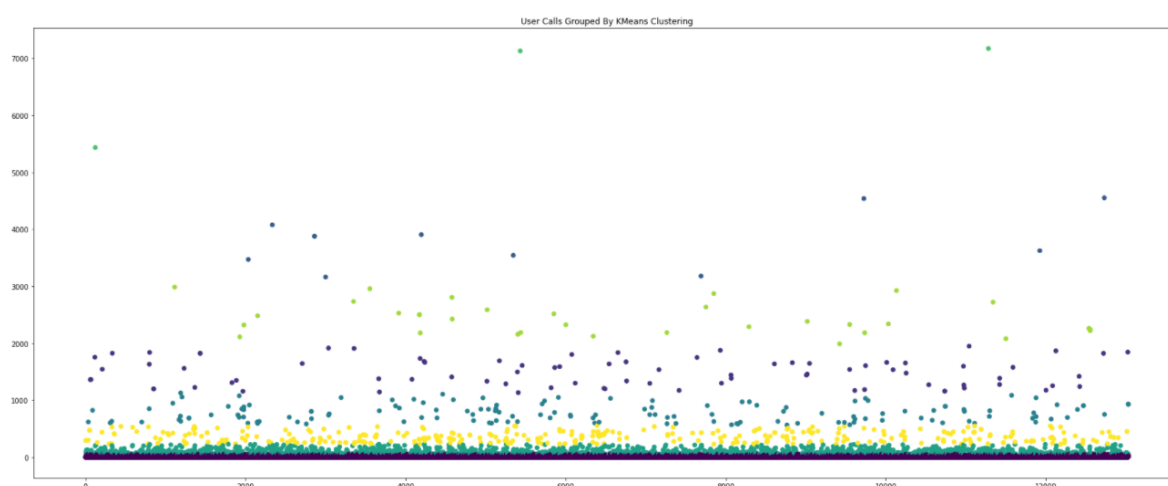
will "rely heavily on data driven machine learning (ML) and artificial intelligence (AI)" [17]. In such networks, this research suggests issues such as outages, as an example, are often the scenario that approaches such as data driven semi-supervised learning is used to address. This research takes inspiration from other researchers who have used similar semi-supervised methods, and other ML approaches such as deep neural networks to create models from call data records [17].

### 3.1. Testing Environment

For the purpose of our investigation, we use Python and the Jupyter notebook environment, both of which are widely used within the data science and machine learning communities. In particular, Jupyter provides an interactive environment that is well suited for repeat experimentation under different parameter configuration, whilst Python has long been established in the field of machine learning, thanks to the wealth of ML libraries that exist such as sci-kit learn, Tensorflow and Keras.

### 3.2. Machine Learning Methods

- **K-Means Clustering:** This unsupervised method of clustering groups similar data points together based on some underlying assumptions about the data. In the case of k-means clustering, the assumption is that  $k$  groups exist, which is either user-specified or may consist of a range of values that are tested for. In Figure 4 we show a simple example of applying k-means clustering to the NoDoBo calls dataset, to establish threshold boundaries between groups of data, where  $k = 8$ .



**Figure 4.** Clustering (k-means) applied to the NoDoBo call duration data

- **Random Forests & Decision Trees:** Using Random (RF) & Decision Trees (DT) can provide predictions using data points from CDR and XDR data. DT can provide regression metrics for call data records by examining features such as user ID or telecom network, then attempt to provide estimated call length or receiver ID. Success of this will primarily depend on the quality of the training stage. With the use RF to mitigate common problems such as overfitting, these predictive methods can be compared to other predictive ML techniques for call data record testing.
- **Linear Regression:** One of the most common methods of machine learning is Linear Regression, easily providing predictive analytics when applied to trained data gathered from existing CDR and XDR data. Tasks such as linear regression can be applied to anomaly detection of call data records, by building a predictive model of expected traffic, and comparing new traffic against this model to estimate the likelihood of anomalous activity. This again, can be compared to other predictive ML methods, attempting to provide higher accuracy when compared.

### 3.3. Applying ML Methods to CDR XDR Data

The primary form of analytic produced from machine learning in the discussed methods is predictive analytics, which can be produced by building a train/test split of existing CDR and XDR data for the purposes of parsing data into a machine learning model.

One study, [8] makes utilization of call data records to model prices of the real estate market, this hybrid ML approach is a highly practical example of call data records used to solve issues whereby modelling of overlapping research sectors is achieved. This application of CDR usage is widespread, and therefore possesses the potential to create ML modelling for other overlapping areas of research. Multi-layer perceptions are trained in this study to build a predictive model, this study follows the pattern of exploring different predictive ML methods that could applied to call data records to produce different types of metrics.

Similarly, [9] remarks to the use of call data records in police investigation. Challenges are posed by big data in this research, various tools and methods are proposed in the effort to assist activities such as crime detection by analysing such records. The process followed in this research involves storing call data records in MapReduce to split files into blocks and distributing them across a cluster [9], whereby this information is passed to a Hadoop file system and processed, due to the large volume of CDR data.

It has been identified, through research that establishes to use a Gaussian mixture model as means of generated model-based metrics for call data records [18]. Gaussian mixture models can provide classification for tasks in call data record analysis that may uniquely provide improved accuracy or prove a more suitable methods than other ML methods which could potentially overlook data points Gaussian methods do not.

## 4. Key Research Challenges In CDR Analysis

In this section, we draw together and discuss what we believe to be the key research challenges in the domain of CDR and XDR data analysis, as identified from the existing literature and from our initial research investigations.

- **Contextualization of Call Activity** Whilst CDR and XDR provides a wealth of information about call records, the classification of calling behaviour of users can quite easily lack contextual information about why behaviours are exhibited (e.g., [19]). Whilst there exist research works that have attempted to use CDR data for predicting house prices [8] and for predicting criminal activity [9], researchers would clearly need to couple multiple data sources together to obtain a suitable level of understanding on the underlying intentions of users.
- **Real time results analysis** Given the sheer volume of information being captured globally, there remains an on-going research challenge into how to manage CDR data in real time (e.g., [20]). Proposed solutions include a distributed approach to data mining of call detail records, however this requires further investigation to understand issues of scalability. Alternatively, a hierarchical approach that provides levels of analysis based on the real-time requirement of the analysis could well be explored in further research.
- **CDR with VoIP** As new technologies continue to evolve, including Voice over IP, and more recently, video conferencing services offered by the likes of Microsoft Teams, Google Meet and Zoom, researchers have noted that the changing distribution function of the duration of calls [21], based on user behavioural changes, will remain as an open research challenge.
- **VoIP Classification Challenges** Focusing on the classification of VoIP traffic U. Anwar et al [22] refers to the issue that presently there may exist a lack of application layer techniques to detect certain protocols and traffic within VoIP. As this research attempts to classify "illegal" traffic, understanding the context of this data is considered a crucial goal.
- **Finding User Habits** Research conducted in the field of data mining of call data records includes an active attempt of identifying user habits Bianchi et al [9] highlights the key summary of



"automatically identifying meaningful information" from meaningful data[23]. Evidence of this can be found in machine learning experiments deriving results in the form of abstract data.

#### 4.1. Future Developing Usage of Mobile Technology

It is important to note, with any exploration and testing of call data records, to understand the shift in mobile technology communication layers moving forward. Generally, the most noteworthy change observed in mobile technology is the advancement of data transfer speeds introduced by improvement to generations.

The data transfer change has been observed with the introduction of fifth generational technology. Supporting 5G as the generational platform for future mobile network development, research such as [24] is an example of research conducted as early as 2018 [24] in an effort to introduce new architecture to fifth generation networks.

The relevancy of call data records could possibly see some diminishment, as a focus is made towards data-intensive application layer technologies, such as those communications that operate on over-the-top application layer mobile applications.

As call data records are a record of transactions made at SMS and traditional calls, many mobile networks may shift call data record analysis tasks to that of VoIP calls that operate at application layers, instead of the 2G supported calls and text messages that CDRs represent. This being considered, a large volume of call data records are still being generated despite these newer, more data-intensive applications being used in preference to traditional telephone calls and SMS.

## 5. Conclusions

It has been identified, through the discussion in this paper, that the use of machine learning models, based on current machine learning tools could be used to create valuable insights into the metrics of call data records. By parsing call data record metrics into a machine learning model, analysts may be capable of performing tasks such as classification of user activity in a more efficient and accurate manner compared to traditional, non machine learning approaches.

As call data records are simply one form of network metric, the opportunity to explore more complex, larger and diverse telecom datasets exists. These challenges have been addressed in such work as Pereira et al [25] which attempts to produce appropriate file system formats for processing XDR that according to this publication and supported study, requires exponential processing power to complete similar operations as to those conducted on traditional call data records.

Additionally, research into modelling human behaviour has also been conducted by means of analysis of call data records Ling et al [26]. This trend of attempting to understand and perhaps predict human behaviour patterns from the analysis of network traffic is also observed in additional research from Li et al [27]. Li et al [27] similarly attempts to model social influences using call data records with the addition of mobile web browsing histories.

Based on the information discussed in this paper, some observations are made into the insight and considerations to be made when processing CDR XDR information for the purposes of machine learning tasks.

Whilst the challenges of processing CDR XDR data are similar to those found in most machine learning scenarios, the observation of context and purpose is observed as a primary characteristic to observe. Understanding how the insights gained from CDR XDR analysis are applied to optimising performance or mitigating threats is an ongoing research challenge.

Future research will address the various emerging techniques to call data records and mobile network data, with the use of machine learning, to create improved predictive metrics and complete further analysis gaining insight to improve mobile network security at an infrastructure level.

**Acknowledgments:** This research was supported by the PhD Partnership Scheme led by the University of the West of England, in partnership with Ribbon Communications Ltd.

## References

1. Bell, S.; McDiarmid, A.; Irvine, J. Nodobo: Mobile Phone as a Software Sensor for Social Network Research. 2011 IEEE 73rd Vehicular Technology Conference (VTC Spring), 2011, pp. 1–5. .
2. Sultan, K.; Ali, H.; Zhang, Z. Call Detail Records Driven Anomaly Detection and Traffic Prediction in Mobile Cellular Networks. *IEEE Access* **2018**, *6*, 41728–41737. doi:10.1109/ACCESS.2018.2859756.
3. Brandusoiu, I.B.; Todorean, G. Applying Principal Component Analysis on Call Detail Records. *Acta Technica Napocensis: Electronica - Telecomunicatii* **2014**, *55*, 25–28.
4. Ruholla Jafari-Marandi, Joshua Denton, A.I.B.K.S..A.K. Optimum profit-driven churn decision making: innovative artificial neural networks in telecom industry. *Neural Computing and Applications* **2020**, *32*, 14929–14962.
5. Songailaitee, M.; Krilavicius, T. IVUS2021: Information Society and University Studies 2021, April 23, 2021, Kaunas, Lithuania. 2021.
6. Leo, Y.; Busson, A.; Sarraute, C.; Fleury, E. Call detail records to characterize usages and mobility events of phone users. *Computer Communications* **2016**, *95*, 43–53. Mobile Traffic Analytics. <https://doi.org/10.1016/j.comcom.2016.05.003>.
7. Rhoads, Daniel.; Serrano, Ivan.; Borge-Holthoefer, Javier.; Solé-Ribalta, Albert. Measuring and mitigating behavioural segregation using Call Detail Records. *EPJ Data Sci.* **2020**, *9*, 5. doi:10.1140/epjds/s13688-020-00222-1.
8. Pinter, G.; Mosavi, A.; Felde, I. Artificial Intelligence for Modeling Real Estate Price Using Call Detail Records and Hybrid Machine Learning Approach. *Entropy* **2020**, *22*. doi:10.3390/e22121421.
9. Kozik, R.; Choraś, M.; Pawlicki, M.; Pawlicka, A.; Warczak, W.; Mazgaj, G. Proposition of Innovative and Scalable Information System for Call Detail Records Analysis and Visualisation. 13th International Conference on Computational Intelligence in Security for Information Systems (CISIS 2020); Herrero, Á.; Cambra, C.; Urda, D.; Sedano, J.; Quintián, H.; Corchado, E., Eds.; Springer International Publishing: Cham, 2021; pp. 174–183.
10. Abba, E.; Aibinu, A.; Alhassan, J. Development of multiple mobile networks call detailed records and its forensic analysis. *Digital Communications and Networks* **2019**, *5*, 256–265. <https://doi.org/10.1016/j.dcan.2019.10.005>.
11. Nair, S.C.; Elayidom, M.S.; Gopalan, S. Impact of CDR data analysis using big data technologies for the public: An analysis. 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS), 2017, pp. 1–6. doi:10.1109/ICACCS.2017.8014651.
12. Lu, Y.; Ma, Y.; Shi, L.; Chen, L. A Deep Learning Approach for M2M Traffic Classification Using Call Detail Records. 2021 IEEE 6th International Conference on Signal and Image Processing (ICSIP), 2021, pp. 964–968. doi:10.1109/ICSIP52628.2021.9688785.
13. Mokhtari, A.; Sadighi, L.; Bahrak, B.; Eshghie, M. Hybrid Model for Anomaly Detection on Call Detail Records by Time Series Forecasting, 2021, [arXiv:cs.LG/2006.04101].
14. Kaur, N.; Ojha, N. Robust fuzzy based clustering approach in data mining using on call data records. 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 2017, pp. 1111–1117. doi:10.1109/ICCONS.2017.8250639.
15. Zhang, S.; Yin, D.; Zhang, Y.; Zhou, W. Computing on Base Station Behavior Using Erlang Measurement and Call Detail Record. *IEEE Transactions on Emerging Topics in Computing* **2015**, *3*, 444–453. <https://doi.org/10.1109/TETC.2015.2389614>.
16. Khaefi, M.R.; Hendrik.; Burra, D.D.; Dianco, R.F.; Alkarisya, D.M.P.; Muztahid, M.R.; Zahara, A.; Hodge, G.; Idzalika, R. Modelling Wealth from Call Detail Records and Survey Data with Machine Learning: Evidence from Papua New Guinea. 2019 IEEE International Conference on Big Data (Big Data), 2019, pp. 2855–2864. doi:10.1109/BigData47090.2019.9005519.
17. Jaffry, S.; Shah, S.T.; Hasan, S.F. Data-Driven Semi-Supervised Anomaly Detection Using Real-World Call Data Record. 2020 IEEE Wireless Communications and Networking Conference Workshops (WCNCW), 2020, pp. 1–6. doi:10.1109/WCNCW48565.2020.9124782.
18. Ficek, M.; Kencl, L. Inter-Call Mobility model: A spatio-temporal refinement of Call Data Records using a Gaussian mixture model. 2012 Proceedings IEEE INFOCOM, 2012, pp. 469–477. <https://doi.org/10.1109/INFOCOM.2012.6195786>.
19. A Gender-Centric Analysis of Calling Behaviour in a Developing Economy Using Call Detail Records.

20. Goergen, D.; Mendiratta, V.; State, R.; Engel, T. Analysis of Large Call Data Records with Big Data. Proceedings of the Conference on Principles, Systems and Applications of IP Telecommunications; Association for Computing Machinery: New York, NY, USA, 2014; IPTComm '14. doi:10.1145/2670386.2670388.
21. Aziz, Z.; Bestak, R. Analysis of Call Detail Records of International Voice Traffic in Mobile Networks. 2018 Tenth International Conference on Ubiquitous and Future Networks (ICUFN), 2018, pp. 475–480. doi:10.1109/ICUFN.2018.8436669.
22. Anwar, U.; Shabbir, G.; Ali, M. Data Analysis and Summarization to Detect Illegal VOIP Traffic with Call Detail Records. *International Journal of Computer Applications* **2014**, *89*. doi:10.5120/15519-2724.
23. Bianchi, F.M.; Rizzi, A.; Sadeghian, A.; Moiso, C. Identifying user habits through data mining on call data records. *Engineering Applications of Artificial Intelligence* **2016**, *54*, 49–61. <https://doi.org/10.1016/j.engappai.2016.05.007>.
24. Rao, R.M.; Fontaine, M.; Veisllari, R. A Reconfigurable Architecture for Packet Based 5G Transport Networks. 2018 IEEE 5G World Forum (5GWF), 2018, pp. 474–477. doi:10.1109/5GWF.2018.8517039.
25. Pereira, O.M.; Capitão, M.; Regateiro, D.D.; Aguiar, R.L.; Osório, J.B. Mediator framework for inserting xDRs into Hadoop. 2016 IEEE Symposium on Computers and Communication (ISCC), 2016, pp. 547–554. doi:10.1109/ISCC.2016.7543795.
26. Ling, F.; Sun, T.; Zhu, X.; Chen, Q.; Tang, X.; Ke, X. Mining travel behaviors of tourists with mobile phone data: A case study in Hainan. 2016, pp. 1524–1529. doi:10.1109/CompComm.2016.7924957.
27. Li, J.Y.; Yeh, M.Y.; Chen, M.S.; Lin, J.H. Modeling social influences from call records and mobile web browsing histories. 2015 IEEE International Conference on Big Data (Big Data), 2015, pp. 1357–1361. doi:10.1109/BigData.2015.7363895.

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.