Finding the Network of Unsocial Elements

Rajesh Kumar Yadav[*], **Archit Gupta**[*], **Areeshaa Parveen**[*]

[*] Department of Computer Science and Engineering, Delhi Technological University, New Delhi, 110042, India

Abstract- Smartphones have become an essential commodity for humans. It is used by a daily person and is used in retrospective units, and that is why today it is not uncommon that in all cases, the first step in resolving a crime is to break down the CDR of the defendant. Today in almost every case of criminal misconduct, the suspect's cell phone testing plays an essential role in detecting wrongdoing. To track down suspects/criminals, intelligence officials need to process a suspect's Call Detail Record (CDR) that has changed designs from professional co-operatives. It is noteworthy that the anti-friendship groups have their own organisation and relationships with different offenders and are opposed to sections of society. They are often linked together by doing evil, or they may be familiar with the abuse. Many of their partners could not be convicted or their names kept in old cases. Writing a CDR to cover the benefit of the knowledge base is a sensible organisation to provide tangible evidence of potential suspects in a variety of contexts. This test paper examines a case-killing case model using CDR.

Index Terms- Graph theory, Call Detail Record

I. Introduction

The purpose of the misinformation is to identify the design and examples that exist between offenders and those that are anti-social. Such an investigation will help deal with many rare cases and provide sensitive information to an organisation that understands the relationship between offenders. Police divisions often come with their own data when fraud, photography, malpractice and many other essential details related to crime are removed[2]. Despite the fact that these structures are very different from office by the organisation, the key objectives and powers are the same. These structures are essential sources of information in police investigations and law enforcement. Especially today, it is an excellent battle for wise and prudent investigators to regain the lead in their office structures. Most of the time, advanced systems develop the skills to create paper reports and get one case at a time. All in all, these modified plans do not give you the option to download much of the information needed for test purposes. There are various sources of information on trial offices that are widely used to investigate a case. One of them is the Call Detail Record (CDR)[8]. A caller ID record (CDR) is a record of information created by trading on the phone or other broadcast communication gear that archives call deception or exchange of broadcast communications (e.g., instant messaging) travelling through that particular office or gadget In this work, we look at a model that shows how the Call Detail Record (CDR) Database, which contains CDR from past cases and hoodlums who have been sued, may be used to research a topic.

II. Problem Statement and Research Objective

Today it is not unexpected that in all cases the primary purpose of dealing with crime is to disband the Call Records of the Suspects. This often requires some investment and effort. Time and complexity of the investigation increase compared to the CDR volume collected. CDR testing has become an integral part of all sin tests. However, it is evident that in most cases, the analytical offices usually collect CDR when preparing the case with the suspects explicitly. They often break it down by looking at specific issues and filing individual suspects. These CDRs are deleted once the case is closed. CDR testing is also recognised as a difficult and time-consuming job, necessitating the use of a product tool. There are other such instruments on the market; nevertheless, as previously said, they can assist in distinguishing CDR from litigation and prosecution. We suggest that CDR linked to older cases and hoodlums can be stored in an integrated database, which may be called a CDR Website, for testing and the reason for testing in different contexts. A CDR-based assessment can be helpful in distinguishing the relationships of habitual offenders or the like, who are involved in a large number of other cases.

III. Proposed model

This paper proposes a diagram based on a comprehension model using a CDR data set. The CDR data set contains CDR for old cases and hoodlums related to those malicious actions. The goal of this research paper is to help the analytical office to investigate cases with the help of CDR. It is noteworthy that the anti-friendly parts have their own organisation and relationships with the various hoodlums and the private parts[1][19]. Often they are related to doing wrong or they may be acquainted with each other's trials and wrong actions. Many of their partners could not be convicted or their names kept in old cases. Assuming that CDR of older cases and suspects are placed in a centralised information base, it is often helpful for testing. Investigating this relationship between the opposing parties can help the analytical office by breaking down many values[12]. However, recognising visitor numbers and their communication on equal information is very difficult. We therefore use graphical expressions, based on the drawing concept. This diagram can show with a picture what the telephone numbers / IMEI are like related to other IMEI calls /numbers. Figure 1 depicts the diagrammatic representation of the proposed model.

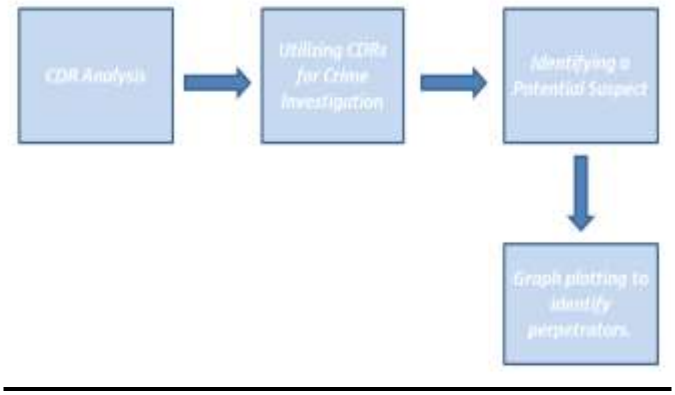
CDR Analysis
Utilizing CDRs for Crime Investigation
Identifying a Potential Suspect
Graph plotting to identify perpetrators.

Figure 1: The flow diagram of the model.

IV. Implementation

This section shows our CDR test model. For representation reasons we used a complete cycle and presented the CDR table in a more precise format, although it actually contains many fields and features that are important for testing. The most essential fields in the CDR are Dialler's Number, Dialler's IMEI number, Caller Number, in addition, Dialler's location and more. For representation reasons we show a few fields in the model. IMEI number is quint essential field of the CDR, which is not mentioned in the framework to reduce the complexity of the diagrammatical representation. IMEI number is essential for identification of items and is vital if the suspect changes the sim cards on the device being used. The subsequent sections discusses the proposed model handling the hidden objects within the model.

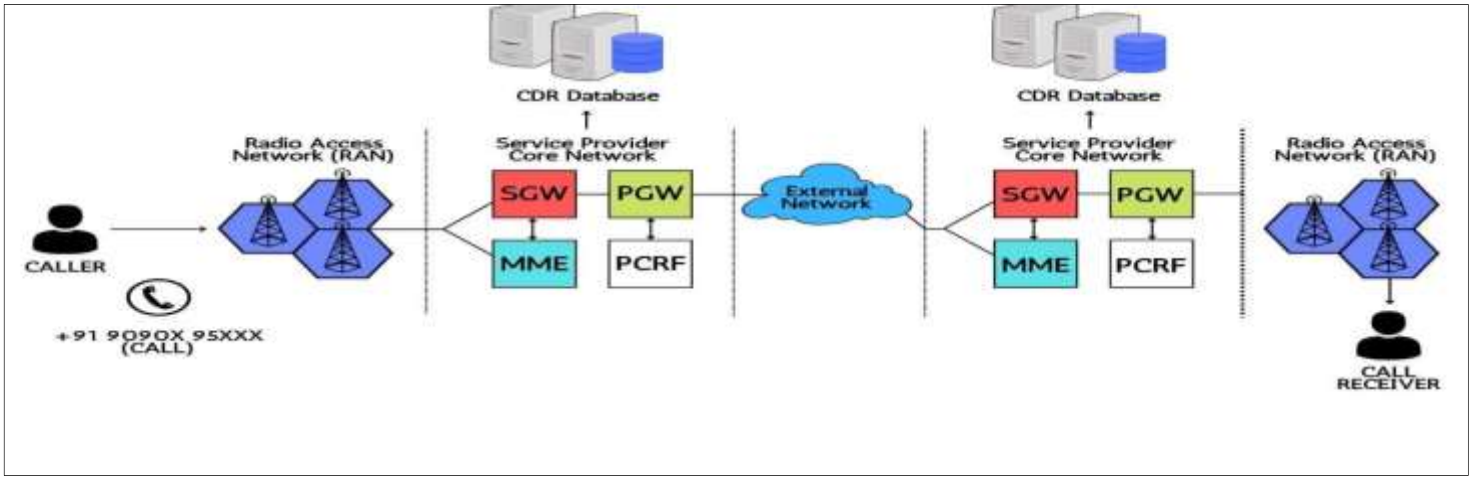
CDR Database
CDR Database
Radio Access Network (RAN)
Service Provider Core Network
SGW
PGW
MME
PCRF
External Network
Service Provider Core Network
SGW
PGW
MME
PCRF
Radio Access Network (RAN)
CALLER
+91 9090X 95XXX (CALL)
CALL RECEIVER

Figure 2: Source of generation of various CDRs for different Telecom providers.

A. CDR Analysis

Today in many negative tests, the Call Record Data test plays an important role. The analytical office collects a record of the details of the calls of the suspects from the portable professional body. A call record (CDR) record is information that is transmitted by telecommunications or a different communication gear that contains call or other communication transactions (e.g., instant messaging) passing through that office or gadget[3]. The call details record contains metadata - information about information - that contains information handles that indicate a specific example of media transfer. It does not include text content. The record has various telephone attributes, for example, duration, length, duration, source number, and purpose number.

As a rule, a call log information showing a specific call may include the following:

- Dialler's contact number
- Receiver's contact number
- Call start time
- Duration of the call
- Billing device contact number
- IMEI number of the devices
- Unique identification number of call
- Call type (connected, missed etc.)
- Mode of communication (Voice, SMS, etc.)
- Any matter of experience and so on.

Modern CDRs (Call Detail Records) are much more informative and details containing numerous attributes including the approximate smartphone GPS location (Latitude and Longitude), cell phone tower connected for the call. These details may vary from one subscription provider to another and is proprietary in nature to protect the privacy of its subscriber base[15]. Management needs and a wide variety of strategies vary from professional engagement to a particular organisation[13]. Call Recording Data provided by professional media transfer organisations to the police is usually subject to an ordinary empty organisation (for example,*.xlsx or *.csv document formatting (See Figure 3). These records are obtained on special requests made by the analytics team, which is followed up with privacy related legal issues removal. This record in the organisation, even for a while, contains horrible records. It is a diagnostic function for the specialist to investigate and make any disturbances [5].

The analytical office takes the help of in-house devices to perform CDR investigations. They often break it down by looking at specific cases and prosecuting individual suspects. If the case is handled, these CDRs are discarded.

We suggest that this CDR for individual cases be included in the data set for long-term testing. The opposition parties have their own organisation and relationships with different and hostile crooks in the public sphere. They are often associated with wrongdoing or, on the other hand, are more likely to be acquainted with abuse. Exploring these relationships between opposing parties can help a professional organisation break down too many values[9]. In any case, distinguishing visitor numbers and their connection to open information is very difficult. Appropriately we use image presentation based on a chart view. The diagram shown may show how phone numbers are associated with other phones. To clarify the whole cycle, in the following area, we accept one situation. The name and number displayed in the model are created.

B. Utilising CDRs for Crime Investigation

To demonstrate the usage of Graph Theory (Social Graph and Identifying Churn Nodes)[14][20], we will take an example scenario. Let's assume a planned robbery occurred at Palika Bazaar, Connaught Place, New Delhi, on February 13, between 1800 hrs to 1900 hrs. To narrow down the search, we get our hands on a potential suspect from location where incident was reported during a specified date, time and location of the crime[10].

C. Identifying a Potential Suspect

In the case mentioned above, the crime investigation department would gather the CDR from Telecom providers for a particular telephone tower situated near the crime scene, here Palika Bazaar[11]. To narrow down the possible suspects, they would investigate the call records within the given time span of one hour, i.e. 1800-1900hrs. We can request this as a simple SQL query such as:

SELECTDialer_Number, Reciever_ Number, Start_Time, Duration WHERE
Tower_Number ='4132'
OR Tower_Number ='4133'
OR Tower_Number = '4134'A
ND Start_Time BETWEEN 18:00 AND 19:00;

The SQL query returns all the entries in the CDRs for all the calls initiated within the given hour span with the three nearest cell phone towers located nearby Paalika Bazaar, Connaught Place, and New Delhi.

Let's assume that the above SQL query returns 10 records, as shown in Table. This narrows down the number of potential perpetrators or suspects to these ten users. In this 4G-5G era and with high connectivity, this number would run into hundreds and thousands and it would be practically impossible to narrow down the search for suspects. This task is no longer humanly possible. This is where our proposed model comes into the picture, where automation can reduce the number of man-hours and help us identify the network of anti-social elements.

TABLE I. SAMPLE CALL DETAIL RECORDS

S. No	Dialer's Number	Reciever's Number	Time (in hrs)	Duration (in mins)
1	99XX6216X1	99XX6315X1	18:15:05	5
2	99XX6216X2	99XX6315X2	18:20:19	6
3	99XX6216X3	99XX6315X3	18:27:37	1
4	99XX6216X4	99XX6315X4	18:29:28	7
5	99XX6216X5	99XX6315X5	18:31:48	2
6	99XX6215X1	99XX6316X1	18:33:26	8
7	99XX6215X2	99XX6316X2	18:40:59	8
8	99XX6215X3	99XX6316X3	18:46:08	4
9	99XX6215X4	99XX6316X4	18:47:16	1
10	99XX6215X5	99XX6316X5	18:55:03	3

D. Graph plotting to identify perpetrators.

We desperately need more tests to find our culprit. The above CDR (Table 1) can be combined with the CDR data set. CDRs for previous cases and linked hoodlums can be found in the CDR database. We can only assume there's a link between these visitors and another scheme involving veteran situations.That would be a nice place to start if we believe we observe such a visitor. The number of suspects can be decreased from thousands to hundreds if a sufficient amount of testing is done[6].Table 2 depicts one such scenario. Table 2 contains lines that allude to visitor numbers that have something to do with the convicted offender. We're currently putting together a CDR of ideas (last 30-50 days). These CDRs can be linked to the CDR database for additional inquiry to see whether they have any ties to the relevant criminals or are part of previous cases[4]. This is our suspects' second point of contact.The visual effects of their association may look as shown in Figure 4:

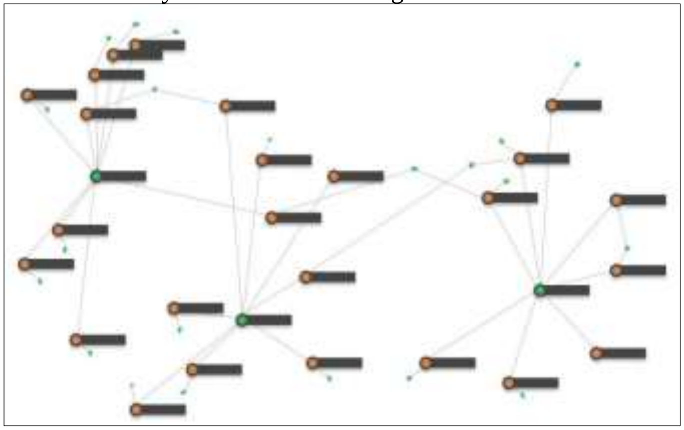

Figure 3: Graph for calls made up to second degree connection of all the probable suspects.

TABLE 2. POSSIBLE SUSPECTS (HIGHLIGHTED ROWS)

S. No	Dialer’s Number	Reciever’s Number	Time (in hrs)	Duration (in mins)
1	99XX6216X1	99XX6316X1	18:15:05	5
2	99XX6216X2	99XX6316X2	18:20:19	6
3	99XX6216X3	99XX6316X3	18:27:37	1
4	99XX6216X4	99XX6316X4	18:29:28	7
5	99XX6216X5	99XX6316X5	18:31:48	2
6	99XX6215X1	99XX6315X1	18:33:26	8
7	99XX6215X2	99XX6315X2	18:40:59	8
8	99XX6215X3	99XX6315X3	18:46:08	4
9	99XX6215X4	99XX6315X4	18:47:16	1
10	99XX6215X5	99XX6315X5	18:55:03	3

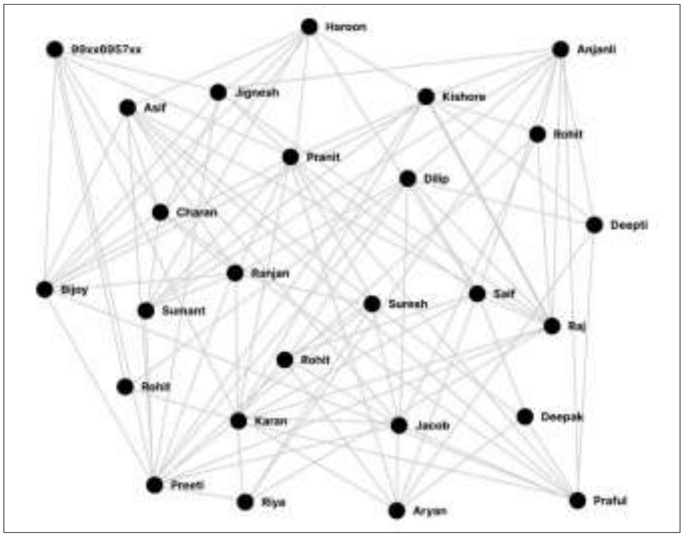

Figure 4: Network of users with anti-social elements.

The calls made by each of our suspicions are shown above. We need to track down the persons our suspicions are calling in order to figure out who they are. Figure 4 depicts the defendants' apparent exposure as well as their interactions with authorities related to the section of the wrongdoing recorded in the police report. As you can see from Figure 4, investigators recognise individual names. It can be implied from the CDR police database acknowledges the offender's name, so compiling a number against a number is a word you can think of. This provides easy access for the observer to separate the connection. The representation of the drawing makes it easy to look for and find related people. The diagram above depicts how our suspects are organised. Remote testing with Excel or normal scientific tools would be required for that data.

V. Using Graph Theory and Churn Prediction

After separating the suspects, the next step is to visually inspect their organisation to zero down the suspect. Currently, in our model, we would find observation of 3 suspects with police history, 99XX6216X2, 99XX6215X4 and 99XX6216X3. Figure 4 shows the organisation of our 3 suspects and a convicted criminal on a police record. The diagram is really thick, and in this way, it is difficult to read and analyse. From the chart above, we can select any suspects to see the accused and convicted organisations, as shown in Figure 5.

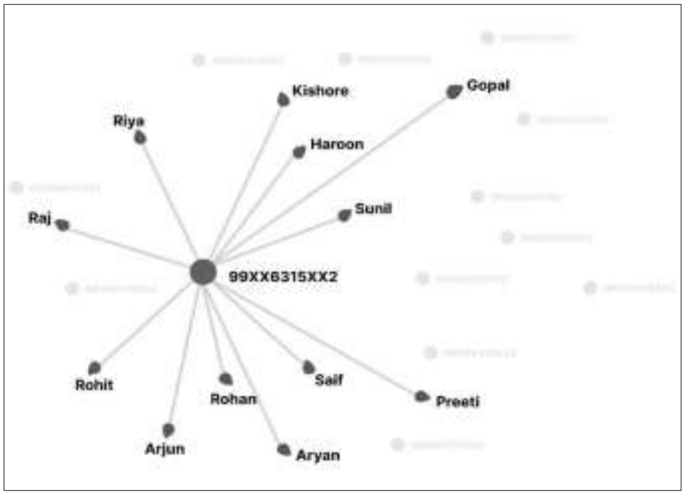

Figure 5: Connections of users with convicted criminals.

After investigating the suspected perpetrator associated with the alleged offenders, the police inspector is unable to identify many of the suspects' suspects without hesitation. The strength and weight of the interactive feature can be determined by the variable dynamics such as the number of calls, traded messages and more. In our model case, suspect 99XX6315X2 is closely related (Figure 6) to two convicted conspirators: Rohan and Arjun The agent also found the criminal history or background of Rohan and Arjun who had been involved in similar crimes. Rohan and Arjun are not involved in the criminal investigation we are investigating. Still, they are in contact with a suspect (for example, 99XX6315X2 about our situation) and possibly a criminal who has done so. We should focus on our testing of him.

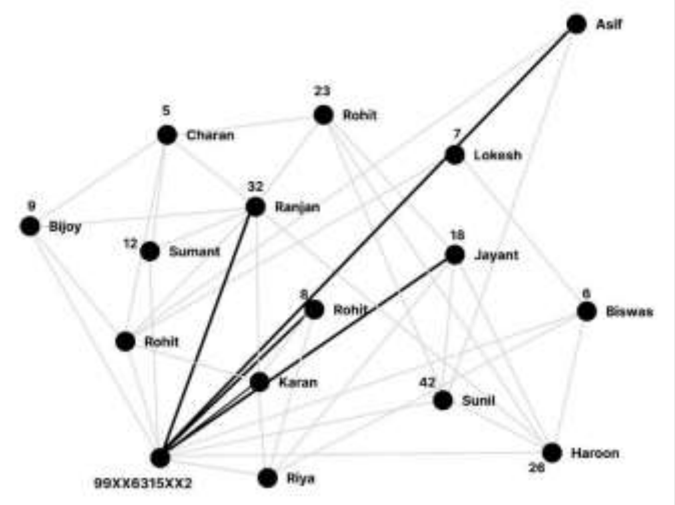

Figure 6: Connection of suspects with anti-social elements depicted using weighted graphs.

VI. CONCLUSION

With the help of our proposed model, not only can we utilise it for crime detection. We can single out suspects to narrow down on the list of perpetrators. To achieve this result, we used the our own to power the charts, which gave us an amazing insight into case-cutting. Negative testing is one of the areas where diagram testing can be used, as well as additional human strategies, to gain experience in complex knowledge. Examining the case using the CDR document may provide additional information if comments are made from the profiles of the organisation of convicted and convicted offenders[17]. As the informal community enjoys more information about the individual, it is often very helpful in testing[16]. We are currently trying to accommodate investigative procedures using CDR next to the Social Network profile appropriately[7].

As a large dataset of CDR is being used hence storing CDR records and their retrieval might be a problem which can be done using hive based retrieval optimization technique[18].

ACKNOWLEDGMENT

Thanks to the contributions of many people, we were able to complete our task and achieve academic success. We would like to express our gratitude to everyone who has helped us. We owe a special debt of gratitude to Dr. Rajesh Kumar Yadav for his indispensable supervision and guidance throughout the project. His motivation and direction have been essential to our development. We also want to thank the department's lecturers for their unwavering encouragement and motivation. Not least of all, we appreciate our parents' support during these tough times.

REFERENCES

[1] F. Ozgul, C. Atzenbeck, A. Celik and Z. Erdem, "Incorporating data sources and methodologies for crime data mining," Intelligence and Security Informatics (ISI), 2011 IEEE International Conference on, Beijing, 2011, pp. 176-180.

[2] Lawrence McClendon and Natarajan Meghanathan, "Using Machine Learning To Analyze Crime Data", Machine Learning and Applications: An International Journal (MLAIJ) Vol.2, No.1, March 2015.

[3] Weiss, G.M. Data Mining in Telecommunications. In Data Mining and Knowledge Discovery Handbook; Springer: Boston, MA, USA, 2005; pp. 1189–1201.

[4] H. Chen, W. Chung, J. J. Xu, G. Wang, Y. Qin and M. Chau, "Crime data mining: a general framework and some examples," in Computer, vol. 37, no. 4, pp. 50-56, April 2004.

[5] Ilona Murynets and Roger PiquerasJover. 2012. Crime scene investigation: SMS spam data analysis. In Proceedings of the 2012 ACM conference on Internet measurement conference (IMC '12). ACM, New York, NY, USA, 441-452.

[6] Scott, J. Social Network Analysis: Developments, Advances, and Prospects. Soc. Netw. Anal. Min. 2011, 1, 21–26.

[7] Scott, J. Social Network Analysis; Sage Publications: Thousand Oaks, CA, USA, 2012.

[8] R. Becker et al., "Clustering anonymized mobile call detail records to find usage groups", Proc. PURBA, pp. 1-8, Jun. 2011.

[9] Richter, Y.; Yom-Tov, E.; Slonim, N. Predicting Customer Churn in Mobile Networks through Analysis of Social Groups. In Proceedings of the SIAM International Conference on Data Mining (SDM 2010), Columbus, OH, USA, 29 Apri–1 May 2010; pp. 732–741.

[10] Wei, C.P.; Chiu, I.T. Turning Telecommunications Call Details to Churn Prediction: A Data Mining Approach.

[11] Nadaf, M.; Kadam, V. Data Mining in Telecommunications. Int. J. Adv. Comput. Theory Eng. 2013, 2, 92–96.

[12] Wasserman, S.; Faust, K. Social Network Analysis: Methods and Applications; Cambridge University Press: Cambridge, UK, 1994.

[13] Hung, S.Y.; Yen, D.C.; Wang, H.Y. Applying Data Mining to Telecom Churn Management. Expert Syst. Appl. 2006, 31, 515–524. [CrossRef]

[14] Amin, A.; Rahim, F.; Ramzan, M.; Anwar, S. A prudent based approach for customer churn prediction. In Proceedings of the International Conf. Beyond Databases, Architectures and Structures (BDAS 2015), Ustron´, Poland, 26–29 May 2015; pp. 320–332.

[15] Amin, A.; Al-Obeidat, F.; Shah, B.; Adnan, A.; Loo, J.; Anwar, S. Customer churn prediction in telecommunication industry using data certainty. J. Bus. Res. 2019, 94, 290–301. [CrossRef]

[16] Yang, C.; Shi, X.; Jie, L.; Han, J. I Know You'll Be Back: Interpretable New User Clustering and Churn Prediction on a Mobile Social Application. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 914–922.

[17] Xevelonakis, E.; Som, P. The impact of social network-based segmentation on customer loyalty in the telecommunication industry. J. Database Mark. Cust. Strategy Manag. 2012, 19, 98–106. [CrossRef]

[18] Droftina, U.; Štular, M.; Košir, A. A diffusion model for churn prediction based on sociometric theory. Adv. Data Anal. Classi. 2015, 9, 341–365. [CrossRef]

[19] Peng, X., Liu, L., & Zhang, L. (2020). A Hive-Based Retrieval Optimization Scheme for Long-Term Storage of Massive Call Detail Records. IEEE Access, 8, 431–444. https://doi.org/10.1109/access.2019.2961692

[20] Dileep, G. K., & Sajeev, G. P. (2021, July 1). A Graph Mining Approach to Detect Sandwich Calls. https://doi.org/10.1109/CONECCT52877.2021.9622627

AUTHORS

First Author – Rajesh Kumar Yadav, Assistant Professor, Delhi Technological University, Delhi, India, 110042,
Second Author – Areeshaa Parveen, Student, Delhi Technological University, Delhi, India, 110042,
Third Author – Archit Gupta, Student, Delhi Technological University, Delhi, India, 110042,

Correspondence Author – Areeshaa Parveen, Student, Delhi Technological University, Delhi, India, 110042.