ANALISIS DAN DETEKSI *FRAUD* PADA DATA PANGGILAN MENGGUNAKAN ALGORITMA K *NEAREST NEIGHBOR* (STUDI KASUS: PT XYZ)

ANALYSIS AND DETECTION OF FRAUD ON CALL DATA USING K NEAREST NEIGHBOR ALGORITHM (CASE STUDY: PT XYZ)

Epta Rizky Alfiyyah[1], Rachmadita Andreswari [2], Edi Sutoyo [3]

[1,2,3]Prodi S1 Sistem Informasi, Fakultas Rekayasa Industri, Universitas Telkom
[1]**rizkyalfiyyah@telkomuniversity.ac.id** [2]**andreswari@telkomuniveristy.ac.id**,
[3]**edisutoyo@telkomuniversity.ac.id**

Abstrak

Penelitian ini membahas mengenai salah satu permasalahan yang terjadi di Indonesia yaitu *telecom fraud.* Salah satu Badan Usaha Milik Negara (BUMN) yaitu PT XYZ, merupakan pihak yang bertanggung jawab dalam menangani perihal *telecom fraud* tersebut. Telecom fraud merupakan suatu aktivitas penggunaan fasilitas telekomunikasi yang dilakukan secara ilegal dan disengaja dalam berbagai bentuk kecurangan, penipuan, atau pun penggelapan oleh orang atau organisasi tertentu yang tujuannya adalah mendapatkan layanan tersebut dan menghindari biaya layanan atau pelacakan rekaman tagihan yang dilakukan secara ilegal. Terdapat beberapa jenis telecom fraud, SIM box fraud merupakan salah satu dari jenis telecom fraud yang akan dibahas pada penelitian ini. Pada penelitian ini dilakukan pengklasifikasian nomor telepon yang terindikasi SIM Box fraud dengan algoritma data mining yaitu K-Nearest Neighbor (KNN). Penelitian dilakukan menggunakan *tool* Jupyter Notebook. Setelah dilakukan pengujian algoritma *K-Nearest Neighbor* dengan menggunakan k=1, k=3, k=5, k=7 dan k=9, maka diperoleh nilai akurasi tertinggi yaitu 74.2% dengan nilai k=9, rasio 0.1, *precision* 74%, *recall* 74%, dan *f1-score* 74%.

Kata kunci : Telecom Fraud, SIM Box fraud, Data Mining, K-Nearest Neighbor.

Abstract

This study discusses one of the problems that occur in Indonesia, namely telecom fraud. One of the State-Owned Enterprises (SOEs), PT XYZ, is the party responsible for dealing with the telecom fraud. Telecom fraud is an activity of using telecommunications facilities that is carried out illegally and intentionally in various forms of fraud, fraud, or embezzlement by certain people or organizations whose purpose is to obtain these services and avoid service fees or tracking records of invoices that are carried out illegally. There are several types of telecom fraud, SIM box fraud is one of the types of telecom fraud that will be discussed in this study. In this study the classification of telephone numbers indicated by SIM Box fraud with the data mining algorithm is K-Nearest Neighbor (KNN). The study was conducted using the Jupyter Notebook tool. After testing the K-Nearest Neighbor algorithm by using k = 1, k = 3, k = 5, k = 7 and k = 9, the highest accuracy value is 74.2% with k = 9, ratio 0.1, precision 74% , recall 74%, and f1-score 74%.

Key words : Telecom Fraud, SIM Box fraud, Data Mining, K-Nearest Neighbor.

1. Pendahuluan

Teknologi informasi atau information technology merupakan teknologi yang menggabungkan komputasi (komputer) dengan jalur komunikasi kecepatan tinggi yang membawa data, suara,

dan video [1]. Teknologi informasi juga berkaitan dengan teknologi komunikasi, dimana teknologi komunikasi menyediakan media penyaluran informasi. Teknologi informasi dan komunikasi dapat diartikan sebagai pemanfaatan perangkat komputer sebagai alat untuk memproses, menyajikan, serta mengelola data dan informasi dengan berbasis pada peralatan komunikasi [2]. Perkembangan teknologi informasi dan komunikasi saat ini salah satunya bisa dilihat oleh perkembangan alat komunikasi. Alat komunikasi adalah alat yang langsung digunakan oleh manusia untuk saling berkomunikasi.

Perkembangan teknologi yang begitu cepat membuat banyaknya teknologi-teknologi canggih yang bermunculan. Salah satunya yaitu munculnya kecanggihan teknologi di bidang telekomunikasi. Telekomunikasi adalah teknik pengiriman atau penyampaian informasi dari satu tempat ke tempat lain [3]. Namun masih ada saja pihak-pihak yang tidak bertanggung jawab yang menyalahgunakan teknologi telekomunikasi ini sebagai bentuk kecurangan. Kecurangan atau fraud menurut Association of Certified Fraud Examiners (ACFE) didefinisikan sebagai "suatu tindakan untuk memperkaya diri sendiri melalui penyalahgunaan secara sengaja atau penggunaan sumber daya organisasi atau aset-asetnya" [4].

Telekomunikasi telah menjadi kebutuhan di seluruh dunia dan telah maju hari demi hari, dimana ini telah mengurangi hambatan komunikasi di seluruh dunia [5]. Tetapi ada saja motif bagi orang yang menghasilkan banyak uang dari mengakses komunikasi secara ilegal dan menggunakannya untuk menghasilkan keuntungan besar, dengan menjual layanan dengan harga jauh lebih rendah dari harga aslinya. Kecurangan telekomunikasi dapat terjadi kapan dan dimana saja, contohnya pada jaringan telekomunikasi, telepon bergerak (mobile communication), perbankan online, dan e-commerce. Menurut Communications Fraud Control Association (CFCA), masalah yang diakibatkan oleh kecurangan dalam telekomunikasi (telecom fraud) ini mewakili hampir $30 miliar secara global pada tahun 2017 lalu [6].

Dikarenakan saat ini banyak terjadi kecurangan di dunia telekomunikasi, dimana kecurangan ini banyak merugikan pihak-pihak yang bersangkutan, yaitu dari pihak penyelenggara telekomunikasi lokal dan operator [7]. SIMBox *fraud* merupakan salah satu jenis dari *telecom fraud* yang ditangani oleh PT XYZ. Dalam penelitian ini akan dilakukan analisis dan pendeteksian SIMBox *fraud* pada data panggilan telepon dikarenakan telah merugikan pihak PT XYZ sebagai pihak yang menangani masalah *fraud* tersebut. Dalam pendeteksian nomor telepon yang terindikasi SIMBox *fraud* dengan menggunakan cara manual yang bukan dengan menerepakan klasifikasi dengan algoritma membutuhkan waktu yang sangat lama sekitar dua jam atau lebih. Maka dari itu pada penelitian ini melakukan analisis dan pendeteksian dengan menggunakan algoritma *K-Nearest Neighbor* dengan alasan *K-Nearest Neighbor* merupakan algoritma *supervised learning* yang mana algoritma tersebut dilatih terlebih dahulu agar dapat melakukan prediksi maupun klasifikasi terhadap data yang akan digunakan. Maka dari itu pada penelitian ini digunakan algoritma KNN untuk analisis dan mendeteksi nomor telepon yang terindikasi sebagai SIM Box fraud dengan judul "Analisis dan Deteksi Sim Box Fraud Di Indonesia Menggunakan Algoritma K-Nearest Neighbor Pada PT XYZ."

2. Dasar Teori

2.1. *Telecom Fraud*

Fraud atau dalam bahasa Indonesia memiliki arti yaitu penipuan. Pada pasal 378 KUHP, yang dimaksud dengan penipuan adalah menguntungkan diri sendiri atau orang lain secara melawan hukum, dengan memakai nama palsu atau martabat palsu, dengan tipu muslihat, ataupun rangkaian kebohongan, menggerakkan orang lain untuk menyerahkan barang sesuatu kepadanya, atau supaya memberi hutang maupun menghapuskan piutang. Telekom fraud merupakan sebuah aktivitas penipuan yang dilakukan pada layanan telekomunikasi dengan tujuan mendapatkan layanan tersebut tanpa harus membayar dan merugikan baik bagi pelanggan dan operator. Penipuan ini hanya bisa dideteksi setelah aktivitasnya telah terjadi, sedangkan pencegahan penipuan adalah tindakan yang dapat digunakan untuk menghentikan terjadinya penipuan tersebut.

2.2. *SIM Box Fraud*

SIM Box fraud adalah penyaluran traffic incoming international yang tidak melalui jalur normal interkoneksi internasional. Namun melalui jalur lain yang pada umumnya menggunakan teknologi VoIP (*Voice Over Internet Protocol*). *SIM Box* merupakan sebuah perangkat yang

mampu memanipulasi nomor ponsel dari pengguna ke penerima. Panggilan dari luar negeri bisa dimanipulasi seolah-olah menjadi panggilan dalam negeri. Secara teknis, sambungan telepon dari luar negeri tersebut diterminasi dan perangkat itu me-*redial* panggilan tadi dan mengalihkannya ke koneksi internet (VoIP) sehingga tidak melalui jalur yang seharusnya dilewati. Ketika panggilan tersebut masuk ke Indonesia, panggilan disalurkan kembali melalui *SIM Box* yang telah berisi kartu prabayar sehingga trafik yang seharusnnya trafik internasional berubah menjadi trafik lokal [8].

2.3. *Call Detail Record* (CDR)

CDR, merupakan istilah yang digunakan dalam sistem *Public Switched Telephone Network* (PSTN) dan *Voice over Internet Protocol* (VoIP), yang merupakan catatan panggilan telepon yang ditulis ke basis data untuk digunakan dalam kegiatan pasca-pemrosesan. CDR mencakup alamat panggilan dan tujuan panggilan, waktu panggilan dimulai dan berakhir, durasi panggilan, waktu hari panggilan dilakukan dan setiap biaya tol yang ditambahkan melalui jaringan atau biaya untuk layanan operator. CDR biasanya digunakan untuk penagihan dan analisis jaringan [9].

2.4. *Big Data*

Big Data didefinisikan sebagai sekumpulan data dengan volume yang sangat besar dan terlalu kompleks untuk dapat diproses dengan teknologi pengolahan data konvensional. *Big Data* juga sering dideskripsikan sebagai data yang memiliki tiga karakteristik, yaitu: *volume, variety*, dan *velocity*. *Big Data* dapat juga didefinisikan sebagai data yang sudah sangat sulit untuk dikoleksi, disimpan, dan dikelola maupun dianalisa dengan menggunakan system database yang biasa karena volumenya yang terus berlipat [10].

2.5. *Data Mining*

Dalam jurnal ilmiah, *data mining* dikenal dengan nama *Knowledge Discovery in Database. Data Mining* merupakan sebuah proses untuk menemukan hubungan, pola dan tren baru yang bermakna dengan menyaring data yang sangat besar, dan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. *Data Mining* mengeksplorasi basis data untuk menemukan pola-pola yang tersembunyi, mencari informasi pemrediksi yang mungkin saja terlupakan oleh para pelaku bisnis karena terletak di luar ekspektasi mereka [11]. *Data mining* secara khusus mencari hubungan dalam kumpulan data yang besar dengan tujuan untuk mendapatkan informasi yang baru. Teknik-teknik *data mining* digunakan untuk menemukan pola yang tersembunyi dan memprediksi tren di masa depan. Keuntungan kompetitif *data mining* yaitu meningkatknya pendapatan, berkurangnya pengeluaran, dan kemampuan pemasaran yang meningkat [12].

2.6. *K-Nearest Neighbor*

K-nearest neighbor (KNN) merupakan kelompok instance-based learning. KNN dilakukan dengan mencari kelompok k objek dalam data training yang paling dekat atau mirip dengan objek pada data atau data testing [13]. Algoritma k-nearest neighbor termasuk ke dalam algoritma klasifikasi. Tujuan dari algoritma klasifikasi adalah untuk memprediksi kelas baru dari dataset yang mempunyai kelas. Menurut Harrington [14], algoritma K-NN banyak digunakan peneliti karena mempunyai kelebihan nilai akurasi yang tinggi, insentive terhadap outlier, dan tidak ada asumsi terhadap data. Tetapi algoritma K-NN juga mempunyai kelemahan, antara lain perlu untuk menentukan nilai k optimal, komputasi yang mahal, dan membutuhkan banyak memori. Rumus dari Euclidean Distance:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2}$$

dimana,

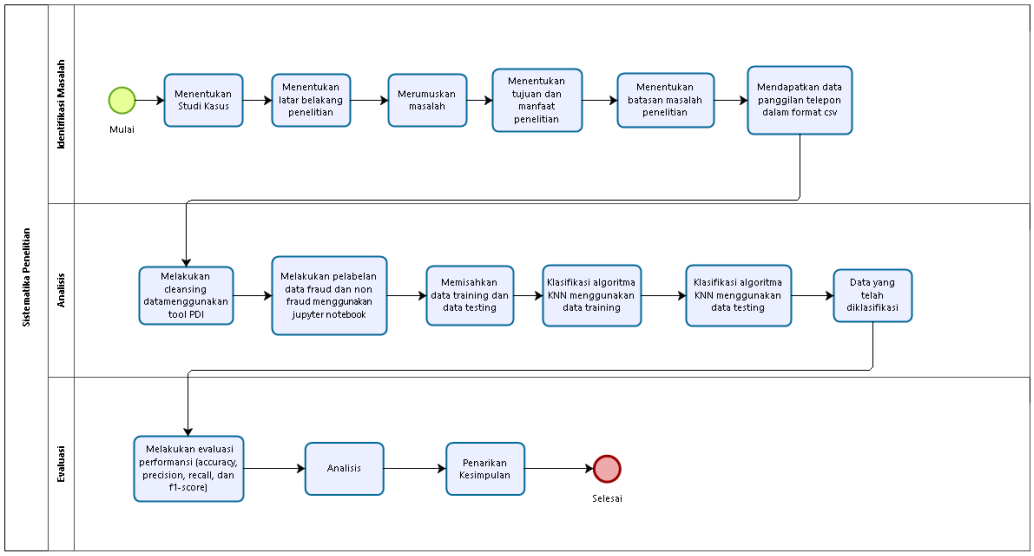
- $\boldsymbol{d(x,y)}$ adalah jarak antara data $\boldsymbol{x}$ ke data $\boldsymbol{y}$
- $\boldsymbol{x_i}$ adalah data *testing* ke-$\boldsymbol{i}$
- $\boldsymbol{y_i}$ adalah data *training* ke-$\boldsymbol{i}$

2.7. Pentaho Data Integration

Pentaho Data Integration (PDI) merupakan sebuah tool open source. Tool Pentaho Data Integration ini menyediakan kapabilitas Extract, Transform, dan Load (ETL) yaitu proses pengambilan, pembersihan, dan penyimpanan data dengan menggunakan format yang sama dan konsisten [15]. Pentaho Data Integration (PDI) bukan hanya untuk ETL saja, tetapi dapat juga digunakan untuk migrasi data ke dalam flat files, cleansing data, dan lain-lain [16]. Setiap proses PDI dibuat dengan menggunakan alat grafis sehingga memungkinkan pengguna menentukan apa yang harus dilakukan tanpa menulis kode untuk menunjukkan bagaimana melakukannya. PDI dapat dikatakan berorientasi pada metadata. PDI mendukung beberapa format input dan output, termasuk text files, data sheets, dan database [17].

2.8. Metodologi

Pada penelitian ini menggunakan metode klasifikasi. Proses klasifikasi ini nantinya akan menghasilkan klasifikasi berdasarkan label yang tercantum pada dataset. Bahasa pemrograman yang digunakan dalam memproses data pra-pemrosesan adalah bahasa Python, sedangkan untuk mengklasifikasikan dataset peneliti menggunakan penambang cepat dan Python berdasarkan perpustakaan Scikit-Learn. Berikut ini adalah sistematika penulisan untuk penelitian ini.



3. Pembahasan

3.1. Dataset

Pada tahap ini dilakukan pengambilan data yang berasal dari Call Detail Record (CDR). CDR adalah data yang berkaitan dengan sebuah penggunaan jaringan telekomunikasi, seperti misalnya panggilan telepon, sms, maupun data service [18]. CDR yang didapatkan yaitu data panggilan telepon yang terjadi pada bulan Juni hingga Agustus tahun 2017. Data yang didapatkan seperti pada Tabel di bawah ini.

Bulan	Total *Rows*
Juni	2138544 rows
Juli	2521822 rows

Agustus	2324961 rows
Total	**6985327 *rows***

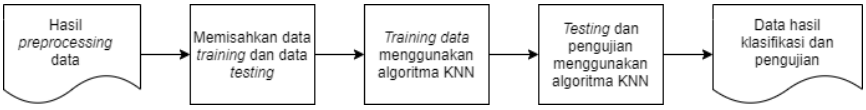
Data panggilan terdiri dari 14 kolom tiap bulannya. Data tersebut merupakan data asli dalam format csv. Dari data yang didapatkan terdapat atribut yaitu Source, Record_id, A_number, B_number, Calling_time, Duration, Call_unit, Ldr, Anouma_duration, Ren_extention, Trunk_in, Trunk_out, Ipaddress, dan Ipaddress_time. Beberapa kolom atau atribut yang tidak digunakan harus dihilangkan sehingga perlu dilakukan cleansing data yang dilakukan pada tool Pentaho Data Integration untuk menghilangkan data yang tidak diperlukan tersebut. Sehingga dari tahap ini diperoleh data bersih yang siap untuk diproses ke tahap selanjutnya. Setelah mendapatkan data, dilakukan pelabelan data menggunakan tool Jupyter Notebook dengan bahasa python. Label yang diberikan terhadap data teridiri dari fraud dan not fraud. Parameter yang digunakan ada tiga, yaitu dari atribut B_Number, Duration, dan Calling_Time. Analisis *labelling data* dilakukan seperti berikut ini:

1. Pada kolom Kesamaan_B_Number dikatakan fraud apabila nomor (B_Number) yang ditelepon oleh A_number selalu berbeda setiap kali melakukan panggilan.
2. Kemudian pada kolom Duration dikatakan fraud apabila durasi panggilan yang dilakukan berada di atas rata-rata. Dimana rata-rata durasi yang didapatkan selama tiga bulan yaitu sebesar 14109 ds.
3. Untuk parameter Calling_Time, dikatakan fraud apabila A_Number melakukan panggilan telepon kepada B_Number secara minimal lima hari berturut-turut.

No.	Nomor	Durasi	Kesamaan_B_Number	Total_Calling_Time	Label
0		16545	0	0	Fraud
1		26775549	8	3	Fraud
2		32616610	23	3	Fraud
3		76472	0	0	Fraud
4		1974	0	0	Not Fraud
5		9242	0	0	Not Fraud
6		107830	5	0	Not Fraud

3.2. Klasifikasi

Klasifikasi pada data mining berfungsi untuk mengelompokkan data berdasarkan keterikatan data terhadap data sampel **[19]**. Pada penelitian ini klasifikasi dilakukan untuk mengetahui seberapa efektif suatu teknik klasifikasi untuk mengklasifikasi target pada Call Detail Record PT XYZ. Pengklasifikasian dengan algoritma KNN ini dilakukan sampai didapatkan Proses klasifikasi dapat dilihat pada gambar di bawah ini.



1. Data hasil preprocessing yang telah dilakukan proses labelling digunakan sebagai inputan.
2. Memisahkan data asli menjadi data training dan data testing dengan perbandingan 90:10, 80:20, dan 75:25. Data training digunakan untuk mempelajari pola data oleh algoritma KNN. Sedangkan data testing digunakan untuk menguji algoritma KNN. Setiap data nantinya diuji dengan menggunakan nilai k=1, k=3, k=5, k=7, dan k=9.
3. Melakukan training data dengan menggunakan algoritma KNN.
4. Melakukan testing data dan pengujian menggunakan algoritma KNN.
5. Menghitung nilai akurasi dari data hasil klasifikasi dan pengujian.

4. Analisis

Selanjutnya adalah menganalisis dan menyimpulkan dari hasil klasifikasi yang telah dilakukan. Dari hasil ini perlu untuk memeriksa keakuratan untuk menentukan kinerja algoritma K-Nearest Neighbor dengan mengevaluasi akurasi, presisi, penarikan dan ukuran F1. Berdasarkan hasil pengujian menggunakan Jupyter Notebook, maka diperoleh accuracy, precision, recall, dan f1-score dari algoritma K-Nearest Neighbor adalah sebagai berikut:

Nilai k	Rasio	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
k = 1	0.1	70.4	70	70	70
	0.2	68.8	69	69	69
	0.25	68.48	68	68	68
k = 3	0.1	72.2	72	72	72
	0.2	71.6	72	72	72
	0.25	70.72	71	71	71
k = 5	0.1	73.2	73	73	73
	0.2	72.7	73	73	73
	0.25	70.48	71	71	70
k = 7	0.1	74.0	74	74	74
	0.2	73.5	73	73	73
	0.25	73.28	73	73	73
k = 9	0.1	74.2	74	74	74
	0.2	73.9	74	74	74
	0.25	73.28	73	73	72

5. Kesimpulan

Dari pengujian pada Call Detail Record yang sudah dilakukan klasifikasi dengan menggunakan algoritma K-Nearest Neighbor, maka diperoleh nilai akurasi tertinggi yaitu 74.2% dengan nilai k=9, rasio 0.1, precision 74%, recall 74%, dan f1-score 74%. Dari hasil pengukuran akurasi, algoritma K-Nearest Neighbor Classifier mampu melakukan prediksi data panggilan telepon yang terindikasi sebagai fraud pada PT XYZ dengan cukup baik.

6. Saran

Diharapkan pada penelitian selanjutnya sebaiknya data yang digunakan untuk melakukan analisis dilakukan pada satu *tool* saja agar lebih efisien dan menghemat waktu. Pembahasan SIMBox *fraud* pada data panggilan telepon selanjutnya bisa dilakukan dengan menggunakan metode lain seperti *clustering* dengan alasan *clustering* bekerja dengan cara mengelompokkan data yang memiliki kemiripan antara satu data dengan data yang lainnya ke dalam masing masing klaster atau kelompok sehingga data dalam satu klaster memiliki tingkat kemiripan yang maksimum dan data antar klaster memiliki kemiripan yang minimum dan juga sebagai bahan perbandingan dengan metode *klasifikasi* pada penelitian ini.

Daftar Pustaka

[1] Williams and Sawyer, "Teknologi Informasi," in *Informations Technology Business Start-Up*, 2019, p. 19.

[2] Y. Maryono and B. P. Istiana, Teknologi Informasi dan Teknologi, Quadra, 2006.

[3] B. Artika, A. C. P. D. Yudanur and V. D. Marlina, 2018. [Online]. Available: https://www.academia.edu/38058912/MAKALAH_SISTEM_TELEKOMUNIKASI.docx.

[4] B. Garner, "Fraud Resources : ACFE," 2004. [Online]. Available: https://www.acfe.com/fraud-101.aspx.

[5] TimKajian, *Naskah Kajian Undang-Undang No.36 Tahun 1999 Tentang Telekomunikasi dikaitkan dengan Perkembangan Konvergensi Telematika*, Jakarta: ftp.unpad.ac.id, 2006.

[6] K. Gonzales, "telecomengine.com," 4 April 2018. [Online]. Available: https://www.telecomengine.com/article/telecom-fraud-29-billion-and-counting-why-it-matters-more-than-ever-in-the-digital-era/.

[7] A. S. Taba, "Ketika Tarif SLI Dimanipulasi," Tempo.co, 2015.

[8] I. Khairuddin, "Begini Cara Kerja Sim Box Fraud," 4 February 2015. [Online].

[9] J. Dong, Network Dictionary, www.javvin.com, 2007.

[10] A. P. Sujana, "Memanfaatkan Big Data untuk Mendeteksi Emosi," *Jurnal Teknik Komputer Unikom*, 2013.

[11] A. M. Siregar and A. Puspabhuana, Data Mining: Pengolahan Data menjadi Informasi dengan Rapid Miner, CV Kekata Group, 2018.

[12] I. K. J. Arta, G. Indrawan and G. R. Dantes, "Data Mining Rekomendasi Calon Mahasiswa Berprestasi di STMIK Denpasar Menggunakan Metode Technique For Others Reference By Similarity To Ideal Solution," *Jurnal Sains dan Teknologi*, 2016.

[13] Wu, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bemotor," 2009.

[14] Harrington, in *Data Mining: ALgoritma dan Implementasi dengan Pemrograman PHP*, 2012, p. 27.

[15] H. Vantara, "Products: Pentaho Data Integration," 20 May 2017. [Online]. Available: https://help.pentaho.com/Documentation/7.1/0D0/Pentaho_Data_Integration.

[16] A. R. Iskandar and I. Intias, "Rancang Bangun Online Analytical Processing (OLAP) Classic Model Data," *Seminar Nasional Informatika dan Aplikasinya (SNIA)*, p. 2, 2019.

[17] A. P. Slavia, "Penerapan Model Autoregressive Integrated Moving Average (Arima) Untuk Prediksi Kemunculan Titik Panas Pada Kabupaten Rokan Hilir," 2019.

[18] IDBigData, 2019. [Online]. Available: https://idbigdata.com/official/explandict/call-detail-record-cdr-analysis/.

[19] I. Oktanisa and A. A. Supianto, "Perbandingan Teknik Klasifikasi Dalam Data Mining Untuk Bank Direct Marketing," *Jurnal Teknologi Informasi dan Ilmu Komputer*, 2018.

[20] BBC, "IBM workers banned from using USB sticks - BBC News," 10 May 2018. [Online]. Available: https://www.bbc.com/news/technology-44069488.

[21] N. Nissim, R. Yahalom and Y. Elovici, *USB-Based Attacks.*, 2017.

[22] B. Payette, Windows Powershell in Action, Shelter Island: Manning Publications Co., 2011.

[23] J. Aiello, D. Coulter, J. P. Jofre and S. , "Getting Started with Windows Powershell," 05 Juni 2017. [Online]. Available: https://docs.microsoft.com/en-us/powershell/scripting/getting-started/getting-started-with-windows-powershell?view=powershell-6.

[24] D. B. Hariyanto, Sistem Operasi, Bandung: Informatika Bandung, 2009.

[25] J. Axelson, USB Complete: The Developer's Guide 5th Edition., Lakeview Research, 2015.

[26] Syngress, "Introduction to Netcat," 2013.

[27] J. Kennedy and M. Satran, "Structure of the Registry - Windows applications | Microsoft Docs," 31 5 2018. [Online]. Available: https://docs.microsoft.com/en-us/windows/desktop/sysinfo/structure-of-the-registry.

[28] H. Wijayanto, "Klasifikasi Batik Menggunakan Metode K-Nearest Neighbor Berdasarkan Gray Level Co-Occurence Matrices (GLCM)," *Seminar Nasional Aplikasi Teknologi Informasi*, 2015.

[29] G. A. J. Satvika, S. Surya Michrandi Nasution and S. M. Ratna Astuti Nugrahaeni, "Penentuan Jalur Kendaraan Terbaik Dengan Klasifikasi Dari Data Mining Twitter Menggunakan K-Nearest Neighbor (KNN)," *Jurnal Universitas Telkom*, 2019.

[30] S. Ghoneim, "Data Science," 2 April 2019. [Online]. Available: https://towardsdatascience.com.