JUIT
ADMI Jurnal Ilmiah Teknik
JUIT Vol 2 No. 2 Mei 2023 | P-ISSN: 2828-6936 E-ISSN: 2828-6901, Page 147-153

STATE OF THE ART FRAUD DETECTION PADA KARTU KREDIT DENGAN MENGGUNAKAN PENDEKATAN ALGORITMA DAN TEKNIK MACHINE LEARNING

Rizki Ariyani
Ilmu Komputer dan Teknologi Informasi, rizkiariyani@staff.gunadarma.ac.id, Universitas Gunadarma

ABSTRACT

This study provides an overview of fraud detection or robbery / fraud detection on credit cards using machine learning as the main technique. This review discusses some research by experts related to the theoretical foundation, advantages and disadvantages, data procedures, analysis methods, and machine learning techniques used. Before big data was widely known and used in the community, fraud detection used the meaning of text and the meaning of data to process the data. Along with the development of technology, many techniques are used in fraud detection, one of which is machine leaning. Machine learning is a branch of Artificial Intelligence that allows computers to have the ability to learn without needing to program anymore. In simple terms machine learning builds an algorithm that allows computer programs to learn and perform tasks on their own without any user features. This kind of algorithm works by building a model from the input or input to be able to produce a prediction or make a decision based on existing data. Machine learning deals with computational statistics that focus on predictions or making based on computer usage. Some of the implementations of machine learning are text analysis, image processing, finance, search and recommendation engines, speech understanding, and so on.

Keywords: *Fraud Detection, Credit Card, Machine Learning*

Abstrak

Studi ini memberikan ulasan tentang state of the art fraud detection atau deteksi pencurian/kecurangan pada kartu kredit dengan menggunakan machine learning sebagai teknik utamanya. Tinjauan ini membahas beberapa penelitian para ahli terkait landasan teoritis, kelebihan dan kekurangan, prosedur pengumpulan data, metode analisis, dan teknik machine learning yang digunakan. Sebelum big data banyak dikenal dan digunakan dimasyarakat fraud detection menngunakan text meaning dan data meaning untuk mengolah datanya. Seiring dengan berkembangnya teknologi, banyak teknik yang digunakan dalam fraud detection salah satunya adalah teknik dengan machine leaning. Machine learning adalah cabang ilmu dari Artificial Intelligence yang memungkinkan komputer memiliki kemampuan untuk belajar tanpa perlu di program lagi. Secara sederhana machine learning membangun sebuah algoritma yang memungkinkan program komputer untuk belajar dan melakukan tugasnya sendiri tanpa adanya instruksi dari penggunanya. Algoritma semacam ini bekerja dengan cara membangun sebuah model dari input atau masukan untuk dapat menghasilkan suatu prediksi atau pengambilan keputusan berdasarkan data yang ada. Machine learning berhubungan dengan computational statistics yang berfokus pada suatu prediksi atau pembuatan keputusan berdasarkan penggunaan komputer. Beberapa implementasi dari machine learning adalah text analysis, image processing, fincance, search dan recommendation engine, speech understanding, dan lain sebagainya.

.

Kata Kunci: *Fraud Detection*, Kartu Kredit, *Machine Learning*

1. PENDAHULUAN

Dalam kehidupan kita sehari-hari, berbagai transaksi dilakukan melalui pembayaran kartu kredit dan transaksi tanpa kartu seperti PayPal, OVO, Gopay, Shoope Pay, dan lain sebagainya. Muncul kekhawatiran yaitu deteksi penipuan atau *fraud detection* yang menyebabkan kerugian besar uang setiap tahun. Jika penipuan berlanjut seperti ini, kabarnya pada tahun 2020 akan mencapai double digit. Saat ini, kehadiran kartu tidak secara fisik diperlukan untuk menyelesaikan pertukaran yang mendorong semakin banyak pertukaran pemerasan [1]. *Fraud detection* adalah salah satu dari dampak ekonomi di era milenial.

Lembaga keuangan harus menggunakan berbagai teknik *fraud detection* untuk mengatasi masalah ini [2, 3]. Terlepas dari semua metode pencegahan yang dilakukan oleh lembaga keuangan dan penguatan hukum dan pemerintah melakukan upaya terbaik untuk memberantas *fraud detection*, *fraud detection* terus meningkat dan tetap menjadi perhatian utama di masyarakat [4, 5]. Kartu kredit umumnya digunakan dalam pengembangan bisnis internet dan juga aplikasi *mobile* dan terutama dalam pertukaran berbasis online. Dengan bantuan kartu kredit, transaksi online dan pembayaran online menjadi lebih mudah dan nyaman untuk digunakan [6, 7]. *Fraud detection* memiliki pengaruh yang besar terhadap perusahaan [8]. Teknik *machine learning* telah banyak digunakan, dan ini menjadi sangat penting di banyak area tempat pengklasifikasi dalam melindungi data. Sistem *fraud detection* mempelajari fitur ekstraksi dan membantu dalam mengontrol *fraud detection*. Penulisan ini bertujuan untuk meninjau literatur tentang penerapan teknik *machine learning* untuk penelitian *fraud detection* pada kartu kredit. Melihat teknik *machine learning* mana yang paling baik digunakan untuk menangani *fraud detection* pada kartu kredit..

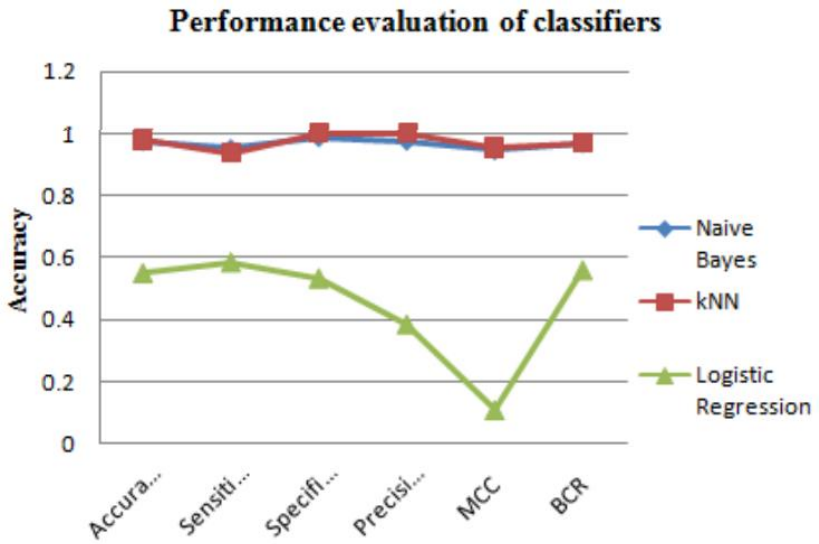
2. TINJAUAN PUSTAKA

Dunia bisnis, perbankan dan *e-commerce* sedang dihadapkan dengan maraknya *fraud detection* yang berkembang beberapa tahun terakhir ini cukup pesat. Khususnya di bidang perbankan, *platform e-commerce*, situs website, aplikasi *mobile*, dan media lainnya. Yang menyebabkan kerugian baik disisi pengguna maupun disisi perusahaan. Sehingga munculah beberapa teknik *machine learning* yang dianggap mampu untuk mengatasi *fraud detection* pada kartu kredit. Dalam penulisan ini menggunakan beberapa perbandingan metodelogi dan masalah yang telah disampaikan oleh beberapa peneliti sebelumnya. Berikut ini adalah acuan perbandingan yang digunakan dalam penulisan ini:

Tabel 1. Perbandingan Penelitian Sebelumnya

No.	Penulis	Topik Penelitian
1.	John O. Awoyemi et al (2017)	Analisis komparatif *fraud detection* pada kartu kredit dengan menggunakan *machine learning*
2.	P. Caroline Cynthia et al (2019)	Analisis komparatif pada pembelajaran *supervised* dan *unsupervised* dengan pendekatan deteksi *outlier* pada *fraud detection* kartu kredit menggunakan *machine learning*
3.	Hasan Najad et al (2020)	*Fraud detection* pada kartu kredit dengan menggunakan *machine learning* dan *deep learning*
4.	Ruttala Sailusha et al (2020)	*Fraud detection* pada kartu kredit dengan menggunakan *machine learning*
5.	Samidha Khatri et al (2020)	Komparasi algoritma *supervised machine learning* pada *fraud detection* kartu kredit

Dalam penelitian pertama masalah yang dibahas oleh John O. Awoyemi et al (2017) penipuan keuangan merupakan ancaman yang terus berkembang dengan berbagai konsekuensi di industri keuangan. *Data mining* telah memainkan peran penting dalam mendeteksi penipuan kartu kredit dalam transaksi online. Deteksi penipuan kartu kredit, yang merupakan masalah data mining, menjadi tantangan karena dua alasan utama. Pertama, profil perilaku normal dan kecurangan yang berubah terus-menerus dan kedua, kumpulan data penipuan kartu kredit sangat miring atau tidak semetris. Kinerja *fraud detection* dalam transaksi kartu kredit sangat dipengaruhi oleh pendekatan sampling pada dataset, pemilihan variabel dan teknik deteksi yang digunakan. Penelitian John O. Awoyemi et al mengkaji kinerja naive bayes, k-nearest neighbour, dan logistic regression pada data penipuan kartu kredit yang sangat tidak simetris. Dataset transaksi kartu kredit bersumber dari pemegang kartu Eropa yang berisi 284.807 transaksi. Teknik *hybrid under-sampling* dan *oversampling* dilakukan pada data tidak simetris. Ketiga teknik tersebut diterapkan pada data mentah dan data yang telah diproses sebelumnya. Pekerjaan diimplementasikan dengan Python. Performa teknik dievaluasi berdasarkan akurasi, sensitivitas, spesifisitas, presisi, koefisien korelasi Matthews, dan rasio klasifikasi seimbang.



Gambar 1. Evaluasi Kinerja Naive Bayes, K-Nearest Neighbor dan Logistic Regression (Sumber: John O. Awoyemi et al, 2017)

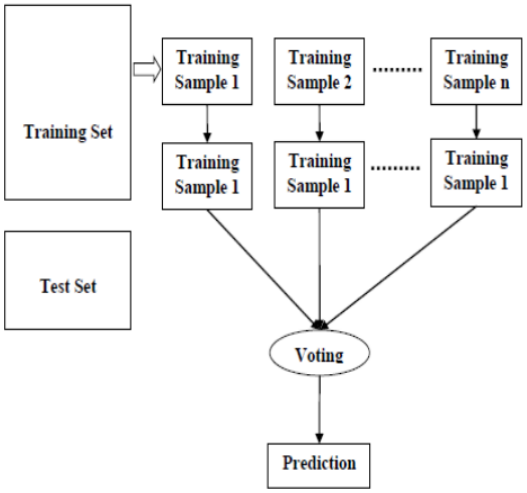
Hasil penelitian John O. Awoyemi et al menunjukkan akurasi optimal untuk pengklasifikasi Naive Bayes sebesar 97.92%, K-Nearest Neighbor sebesar 97.69%, dan Logistic Regression sebesar 54.86%. Hasil komparatif menunjukkan bahwa k-nearest neighbourhood berkinerja lebih baik daripada teknik naive bayes dan logistic regression [9].

Penelitian kedua: P. Caroline Cynthia et al (2019) mengutarakan bahwa penipuan kartu kredit adalah masalah yang relevan secara sosial yang sebagian besar menghadapi banyak masalah etika dan menimbulkan ancaman besar bagi bisnis di seluruh dunia. Untuk mendeteksi transaksi penipuan yang dilakukan oleh pelaku kesalahan, algoritma *machine learning* diterapkan. Tujuan penelitian ini adalah untuk mengidentifikasi algoritma yang paling cocok dan akurat untuk menemukan penipuan atau kecurangan menggunakan algoritma *machine learning* yang diawasi *(supervised)* dan tidak diawasi *(unsupervised)*. Tantangannya terletak pada mengidentifikasi dan memahaminya secara akurat. Dalam penelitian P. Caroline Cynthia et al, pendekatan deteksi outlier dikemukakan untuk menyelesaikan masalah ini menggunakan algoritma *machine learning* yang *supervised* dan *unsupervised*. Keefektifan empat algoritma yang berbeda yaitu *local outlier factor*, *isolation forest*, *support vector machine*, dan *logistic regression* diukur dengan memperoleh skor metrik evaluasi seperti akurasi, presisi, recall score, F1-score, support, dan confusion matrix bersama. Dengan tiga rata-rata berbeda seperti mikro, makro, dan rata-rata tertimbang. Penerapan faktor pencilan lokal memberikan akurasi 99,7 dan hutan isolasi memberikan akurasi 99,6 dalam pembelajaran yang diawasi. Serupa dalam pembelajaran tanpa pengawasan, implementasi mesin vektor dukungan memberikan akurasi 97,2 dan regresi logistik memberikan akurasi 99,8 [10]. Berdasarkan analisis eksperimental, kedua algoritma yang digunakan dalam pembelajaran mesin tanpa pengawasan memperoleh akurasi yang tinggi. Suatu kebaikan secara keseluruhan, serta kinerja yang seimbang, dicapai dalam skor metrik evaluasi dari pembelajaran tanpa pengawasan (*unsupervised*). Dengan demikian, dapat disimpulkan bahwa teknik *machine learning unsupervised* memberikan efisiensi yang sangat baik secara keseluruhan dalam mendeteksi penipuan kartu kredit dengan akurasi tertinggi untuk dataset *real-time* ini. Algoritma ini dapat disimpan di ATM, dan gambar orang tersebut dapat ditangkap. Jika mesin mendeteksi pola yang tidak normal dalam transaksi, maka panggilan peringatan dapat dikirim ke kantor polisi terdekat atau isyarat dapat dikirim ke bank tertentu yang telah menerbitkan kartu kredit tersebut kepada pemegang kartu sebenarnya atau pemegang kartu sebenarnya.

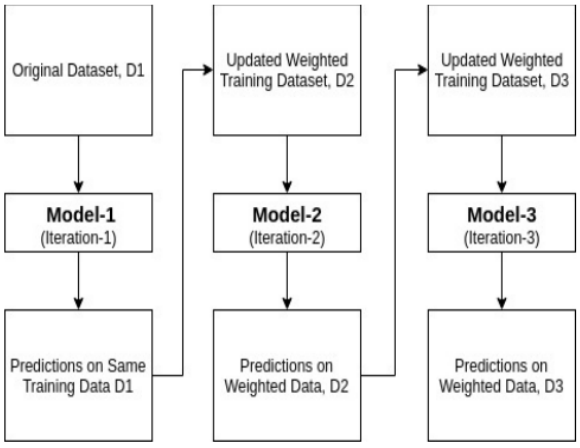
Penelitian ketiga: Hasan Najad et al (2020) kerugian kartu secara global diperkirakan akan melebihi $ 35 miliar pada tahun 2020. Untuk memastikan keamanan pengguna kartu kredit ini, penyedia kartu kredit harus menyediakan layanan untuk melindungi pengguna dari segala risiko yang mungkin mereka hadapi. Karenanya, kami menyajikan pendekatan kami untuk memprediksi transaksi yang sah atau penipuan pada dataset IEEE-CIS Fraud Detection yang disediakan oleh Kaggel [11]. Model penelitian ini adalah BiLSTM-MaxPooling-BiGRUMaxPooling yang didasarkan pada memori jangka pendek dua arah panjang (BiLSTM) dan unit berulang Gated dua arah (BiGRU). Penelitian ini juga menerapkan enam pengklasifikasi pembelajaran mesin yaitu: *Naïve base, Voting, Ada boosting, Random Forest, Decision Tree, and Logistic*

Regression. Membandingkan hasil dari pengklasifikasi pembelajaran mesin dan model kami, hasil menunjukkan bahwa model kami mencapai yang lebih baik karena kami mendapat skor 91,37%.

Penelitian keempat: Ruttala Sailusha et al (2020) mengklasifikasikan transaksi yang memiliki transaksi penipuan dan non-penipuan dalam kumpulan data menggunakan Algoritma *Random Forest* dan Algoritma *Adaboost*. Kemudian kedua algoritma ini dibandingkan untuk memilih algoritma yang paling tepat untuk mendeteksi penipuan transaksi kartu kredit. Alur proses untuk deteksi masalah penipuan kredit pada mencakup pemisahan data, model pelatihan, penerapan model, dan kriteria evaluasi. Dalam model ini mengambil kartu kredit Kaggle yang kumpulan data penipuan dan pra-pemrosesan harus dilakukan untuk data set. Dalam mempersiapkan model harus membagi data ke dalam data pelatihan dan data pengujian menggunakan data pelatihan untuk disiapkan dalam *Random Forest* dan model *Adaboost*. Kemudian mengembangkan kedua model. Akhirnya, akurasi, presisi, *recall*, dan *F1-score* dihitung untuk *bot* model. Akhirnya perbandingan transaksi penipuan kartu kredit lebih akurat. Algoritma *Random Forest* adalah salah satu algoritma pembelajaran yang diawasi secara luas yang dapat digunakan untuk tujuan regresi dan klasifikasi. Tapi, algoritma ini digunakan untuk masalah klasifikasi. Umumnya, Algoritma *Random Forest* menciptakan pohon keputusan di pada contoh data dan mendapatkan prediksi dari masing-masing contoh data. Maka algoritma *Random Forest* adalah metode *ensamble*. Algoritma ini lebih baik daripada pohon keputusan tunggal karena mengurangi *over-fitting* dengan rata-rata hasilnya. Sedangkan Algoritma *Adaboost* digunakan untuk membangun pengklasifikasi yang kuat dari pengklasifikasi yang lebih lemah. Ini dapat dilakukan dengan membangun model yang kuat dengan menggunakan model dalam seri. Awalnya, model dibangun dari pelatihan data. Kemudian model kedua dibangun dari model pertama dengan memperbaiki kesalahan yang mewakili dalam model yang dibuat sebelumnya. Ini adalah proses berulang dan dilanjutkan sampai jumlah maksimum model ditambahkan atau kumpulan data pelatihan diprediksi dengan benar. *Adaboost* adalah salah satu algoritma *boosting* paling sukses yang dikembangkan untuk klasifikasi *biner*. Berikut ini adalah perbandingan Algoritma *Random Forest* dan Algoritma *Adaboost*.

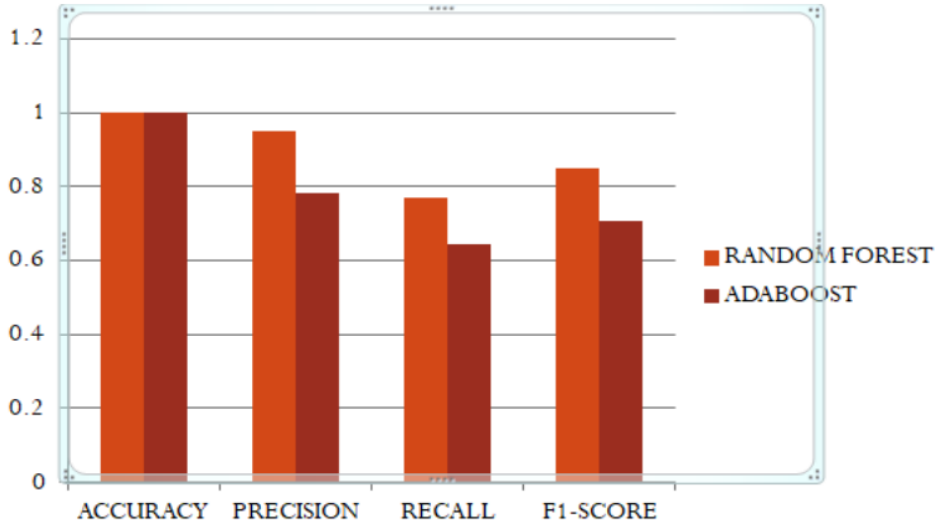


(a)



(b)

Gambar 2. (a) Algoritma *Random Forest* dan (b) Algoritma *Adaboost*.
(Sumber: Ruttala Sailusha et al, 2020)

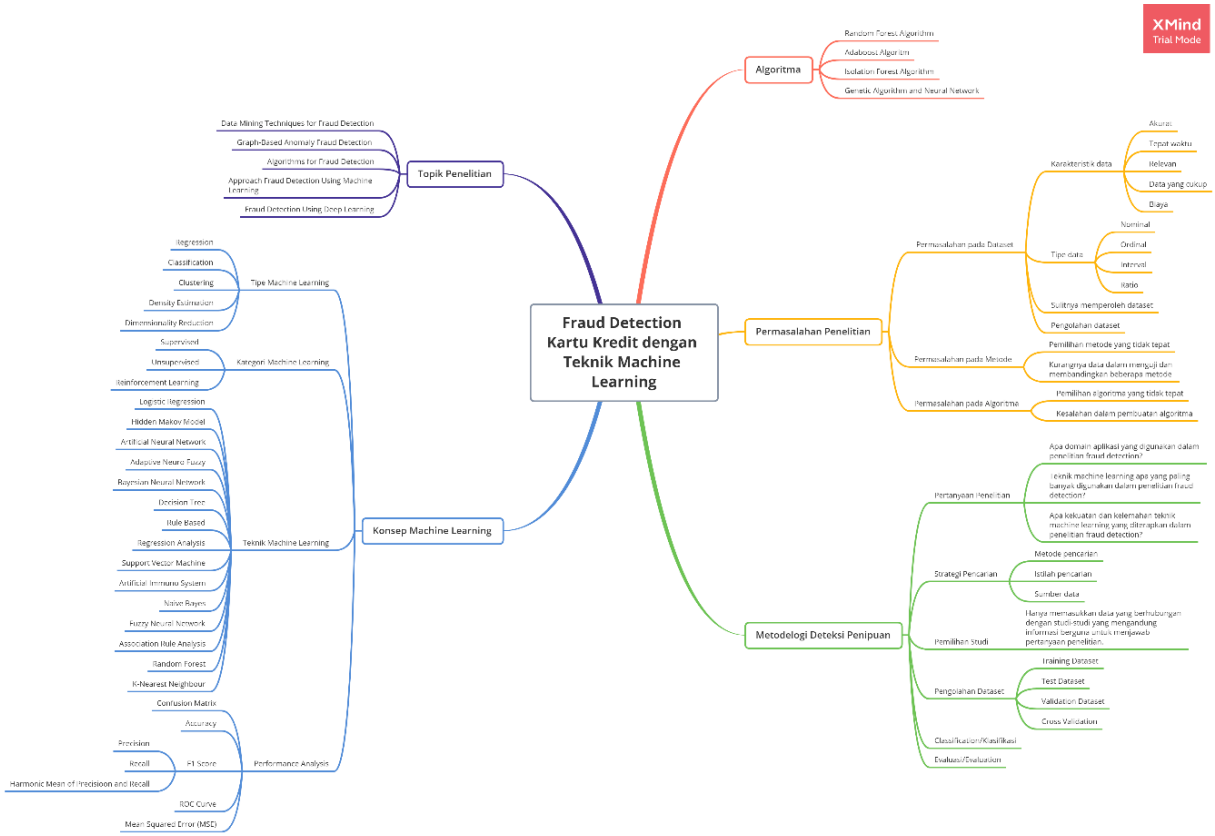


Gambar 3. Perbandingan hasil Algoritma *Random Forest* dan Algoritma *Adaboost*
(Sumber: Ruttala Sailusha et al, 2020)

Dari hasil penelitian kedua algoritma tersebut memiliki akurasi yang sama tetapi presisi, perolehan, dan skor F1 dari kedua algoritma tersebut berbeda. Algoritma *random forest* memiliki presisi, perolehan, dan skor F1 tertinggi dibandingkan dengan algoritma *Adaboost*. Sehingga dapat disimpulkan algoritma *Random Forest* bekerja lebih baik daripada algoritma *Adaboost* dalam hal pendeteksian penipuan kartu kredit [12].

Penelitian kelima: Samidha Khatri (2020) menggunakan data set yang tepat untuk memeriksa kesesuaian berbagai model *machine learning* yang diawasi untuk memprediksi kemungkinan terjadinya penipuan dalam transaksi dengan menggunakan *sensitivity*, presisi waktu sebagai penentu parameter data set. Data set pada dasarnya adalah kumpulan data terkait, dalam tulisan ini menggunakan data set yang tidak diseimbangkan yang tersedia untuk umum. Sebuah Kumpulan data yang tidak diseimbangkan adalah salah satu di mana disparitas terjadi divariabel dependen. Ketidakseimbangan menyiratkan bahwa ada distribusi kelas yang tidak sama. Data set tertentu yang digunakan kumpulan data khusus berisi catatan transaksi yang dilakukan oleh pemegang kartu di Eropa memiliki catatan 284.807 transaksi yang dilakukan selama rentang dua hari yang dimana 492 ditemukan penipuan. Persentase transaksi penipuan diketahui sangat rendah. Set data ini dibuat dan dianalisis lebih lanjut selama upaya dengan menggunakan *Worldline* dan *Machine Learning*. Penelitian ini menggunakan kNN, *Naive Bayes, Decision Tree, Logistic Regression* dan *Random Forest.* kNN adalah salah satu yang paling sederhana tetapi paling model yang efektif, dengan label kelas berupa elemen *data training. Naive Bayes* adalah bentuk model pengklasifikasi probabilistic yang menyiratkan bahwa memiliki kemampuan untuk membuat prediksi untuk beberapa kelas sekaligus. Ini didasarkan pada Teorema *Bayes*. Pengklasifikasi Probabilistik adalah yang memungkinkan untuk memprediksi beberapa kelas data yang memungkinkan untuk memprediksi keputusan dibuat berdasarkan probabilitas bersyarat menggunakan Algoritma tunggal. *Decision Tree* adalah salah satu pemodelan prediktif yang paling banyak digunakan dengan pendekatan dengan model pohon yang terstruktur dalam analisis multi-dimensi dimana terdapat beberapa kelas yang hadir untuk memprediksi nilai keluaran berdasarkan *input* yang disediakan dengan data set.

Dari kelima tinjauan literature yang telah dipaparkan, maka dapat digambarkan *mind mapping* dalam penelitian *fraud detection* pada kartu kredit dengan teknik *machine learning*.



Gambar 4. *Mind Mapping Fraud Detection* Kartu Kredit dengan Teknik *Machine Learning*

Mind mapping tersebut dapat digunakan dalam penelitian selanjutnya untuk kasus *fraud detection* pada kartu kredit dengan teknik *machine learning* dan algoritma yang dapat digunakan sesuai dengan kebutuhan penelitian untuk mendapatkan hasil yang terbaik, terakurat, dan tervalid.

3. KESIMPULAN DAN SARAN

Dari beberapa penelitian yang telah dilakukan oleh para peneliti didapati hasil pengamatan dengan menggunakan teknik *machine learning*. Studi dengan teknik *machine learning* telah berhasil digunakan untuk mempelajari dan menangani *fraud detection* pada kasus kartu kredit dengan hasil yang cukup memuaskan. Pada penelitian pertama hingga penelitian kelima yang telah dipaparkan, terdapat beberapa metode dan teknik *machine learning* yang diangap mampu dalam menangani *fraud detection* pada kasus kartu kredit diantaranya *Naïve base, K-Nearest Neighbor, Voting, Ada boosting, Random Forest, Decision Tree, and Logistic Regression*. Teknik *machine learning* tersebut dapat digunakan untuk mendeteksi penipuan pada kartu kredit yang menghasilkan keakuratan hingga 97%, tergantung pada penggunaan data set, metode, dan algoritma dalam penelitian. *Future work* dari *state of the art* penulisan ini diharapan keamanan dalam penggunaan kartu kredit baik di Indonesia maupun di mancanegara aman dari pencurian dan kecurangan orang maupun pihak yang tidak bertanggung jawab. Algoritma *Machine Learning* digunakan untuk mendeteksi penipuan kita dapat mengamati bahwa hasilnya tidak cukup memuaskan. Jadi, kami ingin menerapkan algoritma *Deep Learning* secara mendalam untuk mendeteksi penipuan kartu kredit secara akurat dan mendetail.

DAFTAR PUSTAKA

[1] U. Fiore, A. De Santis, F. Perla, P. Zanetti, and F. Palmieri, “Using generative adversarial networks for improving classification effectiveness in credit card fraud detection,” *Information Sciences*, Vol. 479, pp. 448–455, 2019

[2] N. Carneiro, G. Figueira, and M. Costa, “A data mining based system for credit-card fraud detection in e-tail,” *Decisions Support System*, Vol. 95, pp. 91–101, 2017

[3] A.C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, "Feature engineering strategies for credit card fraud detection," *Expert System Applications*, Vol. 51, pp. 134–142, 2016
[4] M. Zareapoor, and P. Shamsolmoali, "Application of credit card fraud detection: based on bagging and ensemble classifier," *Procedia Computer Science*, Vol. 48, pp. 679–685. 2015
[5] K. Randhawa, C.K. Loo, M. Seera, C.P. and Lim, A.K. Nandi, "Credit card fraud detection using AdaBoost and majority voting," *IEEE Access* Vol. 6, pp. 14277–14284. 2017
[6] P. Save, P. Tiwarekar, K.N. Jain, and N. Mahyavanshi, "A novel idea for credit card fraud detection using decision tree," *International Journal of Computer Appllication* 161(13), pp. 0975–8887. 2017
[7] S. Sorournejad, Z. Zojaji, R.E. Atani, and A. H. Monadjemi, "A survey of credit card fraud detection techniques: data and technique oriented perspective," ArXiv. 2016
[8] A. Singh, and A. Jain, "Adaptive credit card fraud detection techniques based on feature selection method" *Advances in Computer Communication and Computational Sciences*, pp. 167–178. 2019
[9] J. O. Awoyemi, A. O. Adetunmbi, and S. A. Oluwadare, "Credit card fraud detection using Machine Learning Techniques: A Comparative Analysis," *in 2017 International Conference on Computing Networking and Informatics (ICCNI).* IEEE, pp. 1-9, 2017
[10] P. Caroline Cynthia, and S. Thomas George, "An Outlier Detection Approach on Credit Card Fraud Detection Using Machine Learning: A Comparative Analysis on Supervised and Unsupervised Learning," *Proccedings of ICBDCC 2019 Advances in Intelligent Systems and Computing*, Vol. 1167 pp. 1167, 125-136, 2019
[11] H. Najadat, O. Altiti, A. A. Aquouleh, and M. Younes, "Credit Card Fraud Detection Based on Machine and Deep Learning," *in 2020 11th International Conference on Information and Communication Systems (ICICS),* IEEE, pp. 204-208, 2020
[12] R. Sailusha, V. Gnaneswar, R. Ramesh, and G. R. Rao, "Credit Card Fraud Detection Using Machine Learning," *in Proccedings 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS),* IEEE Xplore pp. 1264-1240, 2020
[13] S. Khatri, A. Arora, and A. P. Agrawal, "Supervised Machine Learning Algorithms for Credit Card Fraud Detection: A Comparison," *2020 10th International Conference on Cloud Computing, Data Science & Engineering (Confluence),* IEEE, pp. 680-683. 2020
[14] Adi Saputra and Suharjito, "Fraud Detection using Machine Learning in e-Commerce," *International Journal of Advanced Computer Science and Applications (IJACSA),* Vol. 10, No. 9, 2019
[15] P. H. Tran, K. P. Tran, T. T. Huong, C. Heuchenne, P. HienTran, and T. M. H. Le, "Real time data-driven approaches for credit card fraud detection," *in Proceedings of the 2018 International Conf.erence on E-Business and Applications.* ACM, pp. 6–9, 2018
[16] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P. Beling, "Deep learning detecting fraud in credit card transactions", *System and Information Engineering Design Symposium (SIEDS)*. IEEE, pp. 129–134, 2018
[17] D. Dighe, S. Patil, and S. Kokate, "Detection of credit card fraud transactions using machine learning algorithms and neural networks: A comparative study," *in 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA).* IEEE, pp. 1-6, 2018
[18] M. Puh and L. Brkic, "Detecting credit card fraud using selected machine learning algorithms," *in 2019 42nd International Converence on Information and Communication Technology, Electronics and Microelectro (MIPRO).* IEEE, pp. 1250–1255, 2019