End of Module Assignment
Launching into Computer Science

Word count: 2081

AI in Financial Fraud Detection: Danske Bank Case Study

1. Introduction

Kute et al. (2021) argue that financial fraud has become a global problem for several decades and is considered one of the biggest threats to the economy and society. This affects various industries, including finance, resulting in significant financial losses and compromising confidence and security. Gee and Button (2019) claim that the annual aggregate volume of money fraud transactions is approximately 6.05% of the global gross domestic product. The ABA Banking Journal (A.B.J) (2022) reports that the scale of losses due to fraud in the financial services sector is huge, and every dollar spent on fraud causes $4.36 in associated costs. Fraud detection and prevention through traditional rule-based and manual processes cannot address the complexity and enormity of modern fraudulent activities. Artificial intelligence (AI) and machine learning (ML) have been established as key technologies for enhancing fraud detection and prevention systems. Recently, the financial sector has seen several high-profile bank frauds. Consequently, anti-money laundering (AML) has become increasingly important for banks and plays a central role in their work. This essay aims to present the main ideas of the most common and most practiced AI techniques and ML models of fraud detection based on the case of the Danske Bank.

2. AI Techniques for Fraud Detection

Traditional rule-based fraud prevention approaches have relied on human-written rules, in the form of "if/then" logic and respond when a predetermined condition is breached (Planque, 2017). If a transaction is flagged as suspicious, the system either rejects the transaction or alerts the team for manual review. For example, a bank's fraud detection system flags supposed clients who use credit cards in odd areas or increase their payment frequency. This method is effective, but labour-intensive and expensive due to numerous rules and keeping a consistent stream of notifications adds to the complexity (Intellias, 2022, Manchev, 2021).

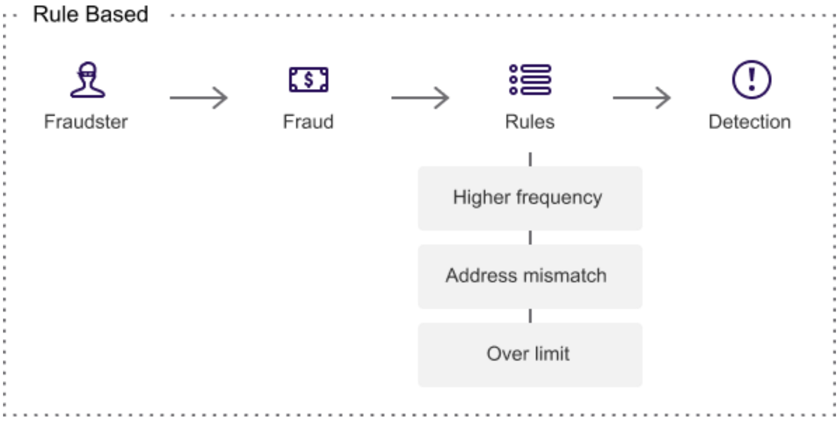

Figure 1. Rule based fraud detection

As far as data analysis is concerned, a computer outperforms a human being, and AI offer huge opportunities (Eliaçık, 2022). Through that they allow us to explore complex patterns humans cannot understand. ML models automatically detect and apply more complex and dynamic rules than traditional systems. It evaluates data from historical fraud incidents, detects patterns and links between data points, and trains models to recognise similar patterns when they appear in future datasets. ML models may predict criminal activities by detecting anomalies, which are subtle and atypical behavioural patterns that turn aside from the norm and could be indicators of upcoming fraud. For example, an e-commerce platform's fraud detection algorithm detects a suspicious credit card transaction that does not match its users' behavioural patterns based on a variety of parameters, such as the visited pages before placing an order.

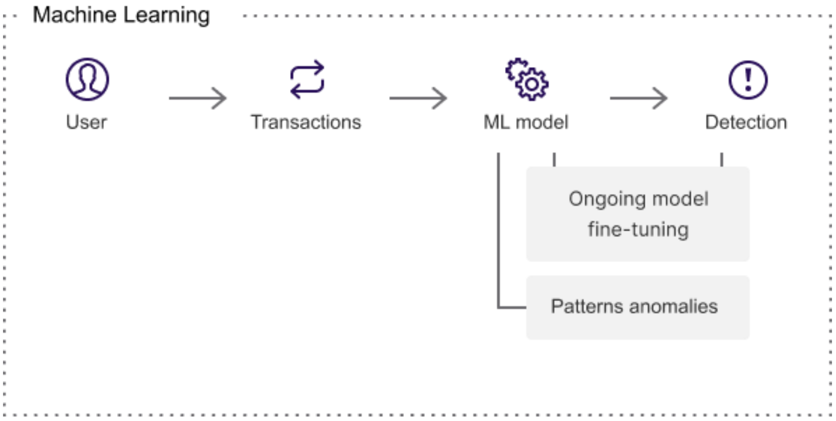

Figure 2. Machine learning fraud detection

These models improve over time, analysing new data, including unmapped data points. If new fraud scenarios emerge, machine learning-based anomaly detection systems swiftly adapt to them, automatically integrating and updating current rules without the need for human interaction. Below is provided a summary of traditional versus ML-based fraud detection:

Table 1. Traditional and ML fraud detection comparison.

Traditional fraud detection	**ML-based fraud detection**
Rule-based systems	Machine Learning and deep learning
Could become outdated as fraudsters develop new techniques	ML models continuously adapt to new fraud patterns
Requires manual updates to rules	ML models learn in real-time and automatically adjust their algorithms
May generate false positives or false negatives	AI models enhance accuracy by considering a multitude of factors. They reduce false positives
Latency issues, especially as transaction volumes increase. Challenging real-time fraud detection.	AI excels in real-time analysis, enabling swift identification of suspicious patterns and immediate response to potential fraud

Alkhalili et al. (2021) divide machine learning into three categories: supervised learning, unsupervised learning, and reinforcement learning. Supervised learning involves human supervision (Shetty et al., 2022), as well as a set of input qualities and an output value. The algorithm learns the link between input and output based on past data. The system can also increase its accuracy by learning from fresh incoming data.

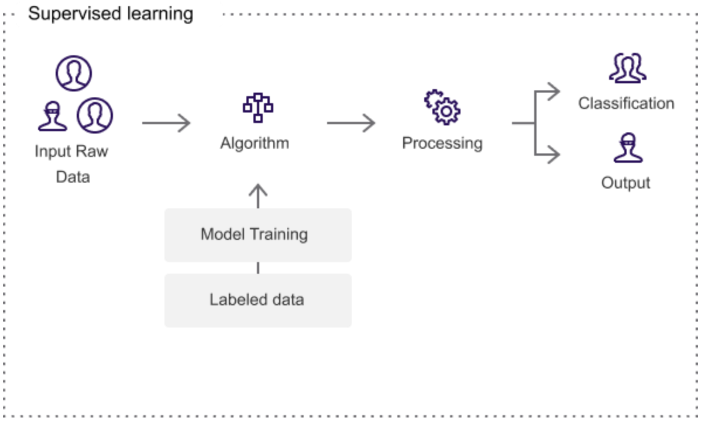

Figure 3. Supervised machine learning

Unsupervised learning is a technique that uses a collection of input qualities but produces no output value. The approach's purpose is to examine the input qualities to identify a pattern of similarity among the data and then categorise it.

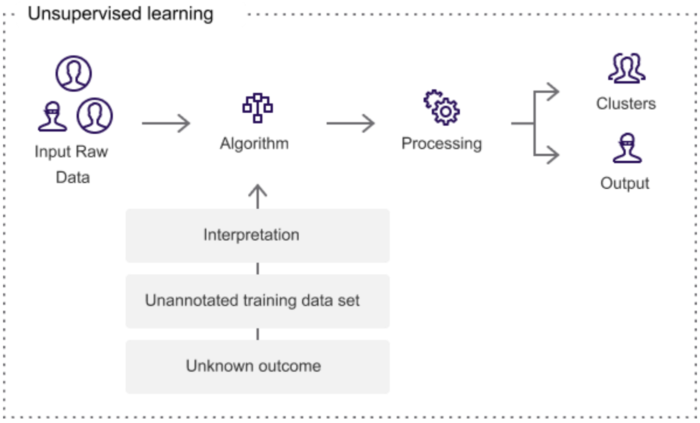

Figure 4. Unsupervised machine learning

Reinforcement learning is a different learning strategy from other methods. It aims to interact with the environment and receive an outcome for each action.

Table 2. Machine learning approaches.

Supervised learning	ML-based fraud detection systems are trained with large amounts of labeled data, previously annotated with certain labels describing its key features. This can be data from legitimate and fraudulent transactions described with "fraud" or "non-fraud" labels, respectively. These labeled datasets, which require rather time-consuming manual tagging, provide the system with both the input (transaction data) and the desired output (groups of classified examples), allowing algorithms to identify which patterns and relationships connect them and apply such findings to classify future cases.
Unsupervised learning	These algorithms are fueled with unlabeled transaction data and have to autonomously group these transactions into different clusters based on their similarities (shared behavioral patterns) and differences (typical vs unusual patterns which can correspond to fraudulent activity). This approach, typically associated with deep learning, is computationally demanding but can be the only choice when facing fraud attempts that have never been met before and therefore unlabeled.
Reinforcement learning	This trial-and-error approach involves multiple training iterations in which the algorithm performs a fraud detection task in different ways several times until it can accurately identify fraudulent and non-fraudulent attempts. Since it does not require labeled inputs, reinforcement learning can be applied without prior knowledge of the current fraud scenario. However, it requires considerable computing power.

Because of its capacity to examine massive volumes of data, discover patterns that might point to fraudulent behaviour and adapt to new data, ML is becoming increasingly popular for fraud prevention and detection (Flavián et al., 2021). By leveraging historical transaction data, customer profiles, and other relevant information, these algorithms can distinguish between legitimate and fraudulent transactions with remarkable accuracy. Banks can identify fraud in real time or before it occurs, thereby minimizing losses and protecting customers. The following are some popular techniques of ML in fraud prevention:

- **Anomaly detection**. Financial transaction data may be analysed using ML algorithms to detect unexpected trends or deviations from usual behaviour (Barnard & Stryker, 2023). The algorithms train using historical data to recognise justifiable transactions and identify suspicious activities that may suggest fraud. It can be classified into three types: a) point; b) contextual, and c) collective anomalies (GfG, 2023).

- **Risk scoring**. ML algorithms may rate transactions or accounts of clients by risk based on based on a variety of criteria (historical data, transactional frequency, etc.) (AMLYZE, 2023). High-risk ratings imply a greater chance of fraud, allowing compliance officers to concentrate on specific transactions or accounts that require additional investigation.
- **Network analysis**. Fraudsters frequently develop networks so they can carry out their schemes and activities. ML techniques, such as graph analysis evaluate relationships between subjects and detect clusters or anomalous relationships (Burke, 2022).
- **Identity verification**. ML algorithms can verify submitted data, like ID documents or face recognition, to confirm a person's identification, as well as to avoid identity stealing (Lieberwitz, 2023).
- **Text analysis**. ML algorithms may automatically analyse meaningful content from unstructured text data (emails, online data and documents) to detect patterns or phrases that may alarm fraud or scams (Miroshnyk, 2022).
- **Adaptive learning**. ML algorithms may retrain and keep up-to-date using new data to be better suited to detect changed and renewed fraud tactics by fraudulent actors (Odmark, 2023).

Identifying a suitable algorithm or algorithm ensemble to implement data analysis for fraud detection may be a complicated task because their performance is influenced by the circumstances in which financial institutions operate. Several ML algorithms were selected from the list published by Fraud.com (2023), which sorts algorithms by academic publication frequency, as well as combined with use cases for fraud detection.

Table 3. Machine learning algorithms.

Logistic regression	A supervised learning algorithm which calculates the probability of one event out of two alternatives, such as "fraud" and "non-fraud" based on a set of relevant parameters.
Decision tree	Another algorithm of the supervised learning subset is a tree-like decision-making model in which every bifurcation represents the analysis of a certain metric or condition (spending threshold, location, etc.) to determine whether an operation is fraudulent.
Random forest	A combination of several decision trees to further expand the amount of data types and conditions examined and identify non-linear relations among multiple variables.
Support vector machine	Several systems rely on this supervised learning algorithm for credit card fraud detection because of its excellent performance with large datasets despite it being computationally demanding.
K-nearest neighbor	This supervised learning algorithm, which proves quite accurate but difficult to interpret when making a certain decision, can frame the nature of an event (be it fraud or non-fraud) by comparing it with similar occurrences recorded in the past.
Neural networks	Complex, multi-layered architecture and superior big data analysis capabilities make neural networks the go-to algorithms when it comes to spotting non-linear relations and dealing with unprecedented fraud scenarios through supervised, unsupervised and reinforcement learning.

ML provides a better rate compared to conventional manual fraud detection methods regarding speed, accuracy, and costs. ML systems can process new data automatically, and continuously update detection models in real-time without constant human supervision. It is faster and more accurate fraud detection, relatively reducing false positives and negatives as compared to the traditional approaches. Secondly, ML algorithms can utilize data 24/7 without compromising on accuracy, which can be deployed for continuous fraud detection.

Despite its benefits, ML-based fraud detection should be supported by continuous monitoring by human experts to improve and optimise the performance of ML models to ensure comprehensive fraud protection. Furthermore, the dark side of AI and ML should be recognized because fraudsters are also using these technologies to develop ever-more sophisticated fraud techniques.

3. Danske Bank and AI fraud detection

The Danske Bank (DB) is one of the largest banks in the Nordic region, operating for 152 years (Forbes, N.D.) and providing a range of banking services (retail, corporate banking, asset management, and investment banking). The financial institution is present in 16 European countries, as well as in North America and India.

However, DB found itself to be a part of a significant scandal, called the DB Money Laundering Scandal or DB Estonia Scandal. This crisis erupted when a significant illegal financial movement of approximately €200 billion in suspicious funds was revealed within the bank's Estonian branch from 2007 to 2015 (U.S. Department of Justice, 2022). DB faced significant legal and regulatory consequences as a result of the scandal. It incurred hefty fines, including a record of €2 billion fine imposed by the US and Danish authorities in May 2020 (Milne, 2022; U.S. Department of Justice, 2022). It revealed weaknesses in banks' governance, risk management practices, compliance procedures, and wider discussions on improving AML frameworks in the banking industry (Gottschalk, 2022). As a result, banks have ceased operations in Russia and Latvia, focusing on its core Nordic market (Biscevic, 2019) while maintaining a service center in Lithuania (LRT, 2019). According to Bloomberg reporter Schwartzkopff (2021), the CEO of DB Carsten Egeriis confirmed the credibility and reputation of the scandal-damaged DB and reduced investor confidence.

To address fraud challenges, DB in partnership with Teradata Consulting, moved towards a data-driven approach (Kureishy et al., 2018). This involved reducing false-positive and increasing true-positive fraud-detection rates. As shown in Figure 4, DB rely on large-scale servers with an expert-driven approach, typically involving a rule-based fraud engine. The left-hand side of the system is the existing setup of payment infrastructure, based on intuition and some light analysis of business-created rules, running with almost no changes. The infrastructure secured a large number of transactions from fraud; however, its high false-positive rate was relatively costly, and updating and maintaining it as fraudsters developed was impractical. In addition, the bank understood that fraud would worsen in the short and the long run because of the rapid digitalization of the banking system, and therefore, it recognizes that new technologies need to be used ahead of fraudsters.

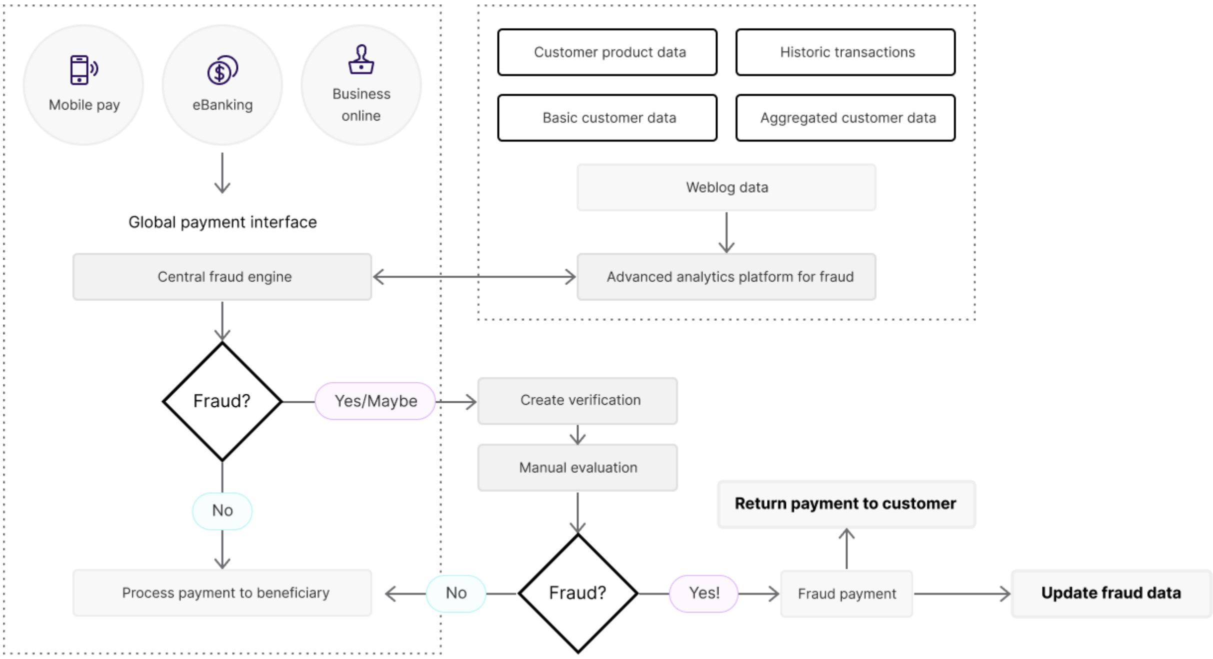

Figure 5. Danske Bank's fraud detection system operation.

The DB's left-side fraud detection engine was enriched with Deep Learning (DL) algorithms, which allow for the analysis of tens of thousands of latent features by using the same data to feed into the right-hand side. An ensemble of DL models was tested, as follows:

- A Convolutional Neural Network (ConvNet) algorithm was used for visual data analysis, taking images as an input and assigning importance to different objects, extracting features, and identifying patterns in the image (Torabi et al., 2023; Chen & Lai, 2021). It helps the system to learn from temporary and static information to acquire insight into the characteristics of fraud by converting transactions into a 2D picture.

- Long Short-Term Memory (LSTM) networks are a form of the recurrent neural network, that deals with algorithms that attempt to replicate the human brain's operation and find underlying links in sequential data (Subrahmannian, 2023; Khusheef et al., 2022). In other words, it learns from temporal information and classifies whether a sequence of transactions contains fraud.
- Autoencoders are models that minimise the amount of input data by reproducing it (Singh, 2024). These models have emerged as effective techniques to identify anomalies. Therefore, training a neural network to reproduce transactions that look like non-fraud data does a dimensional reduction of the reconstruction error rate for fraud cases and blows it up again to try to reproduce the transaction (Dzakiyullah, 2021).

Figure 6 shows the base of the rule-based engine, the increase in efficiency associated with classic ML, and the much greater lift resulting from the collection of deep learning models.

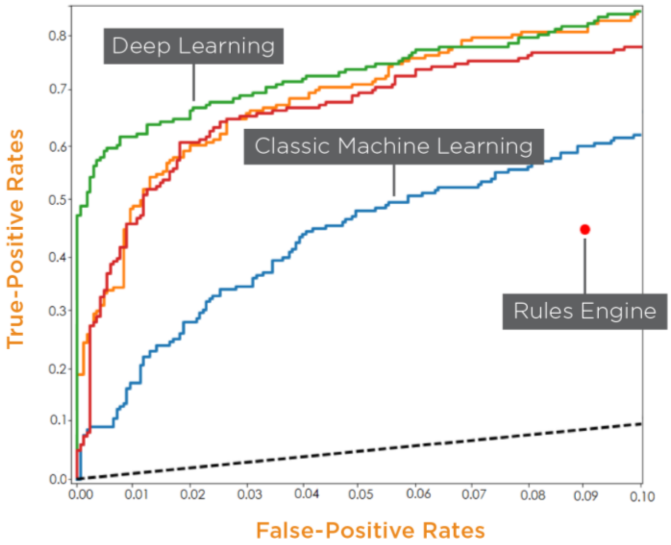

Figure 6. Comparison of results of DL models, classic ML, and legacy rule-based engines (Kureishy et al., 2018).

After testing several DL models, such as ConvNet, LSTM, and autoencoder, the residual network (ResNet) (a type of convolutional neural network and an algorithm commonly used for object identification and computer vision) was found to be very effective.

4. Conclusions

In conclusion, the adoption of AI and ML technologies in financial fraud detection, as can be seen in the case of DB, is becoming a common practice, and it is the most efficient method in combating rampant fraud in the financial industry.

ML-based anti-fraud measures can ensure multidirectional protection with anomaly detection, risk scoring, network analysis, identity verification, text analysis, and adaptive learning. These techniques exploit historical transaction data, customer profiles, and other related information to differentiate between real and fraudulent transactions with shocking accuracy, thereby enabling banks to mitigate losses and safeguard customer assets in real time.

The cooperation between DB and Teradata Consulting exemplifies the transformation to a data-driven approach involving deep learning algorithms such as ConvNets, LSTMs, and Autoencoders in its fraud detection system. These cutting-edge algorithms allow the analysis of latent features, time series, and dimensional reduction, among others, and can be used to detect fraud patterns and identify anomalies within the bank's operations.

The DB case features the necessity for financial institutions to employ AI and ML technologies to continue to change fraud schemes and protect their integrity and image in the increasingly digital environment of the banking sector. Through the application of AI-driven fraud detection tools, banks will be able to strengthen their defenses, improve business efficiency, and engender trust among customers and investors.

Reference list

A.B.J. (2022) Survey finds fraud costs rising for Banks, *ABA Banking Journal*. Available at: https://bankingjournal.aba.com/2022/11/survey-finds-fraud-costs-rising-for-banks/ [Accessed: 11 February 2024].

AMLYZE (2023) *Managing AML risk assessment: Tools for customer evaluation*, *Amlyze*. Available at: https://amlyze.com/aml-risk-assessment/ [Accessed: 6 February 2024].

Barnard, J. & Stryker, C. (2023) *What is anomaly detection?*, *IBM*. Available at: https://www.ibm.com/topics/anomaly-detection [Accessed: 7 February 2024].

Biscevic, T. (2019) *Danske Bank to close Estonian Branch, eba to investigate regulators*, *OCCRP*. Available at: https://www.occrp.org/en/daily/9268-danske-bank-to-close-estonian-branch-eba-to-investigate-regulators [Accessed: 12 January 2024].

Burke, J. (2022) *What is the role of machine learning in networking?: TechTarget*, *Networking*. Available at: https://www.techtarget.com/searchnetworking/answer/What-is-the-role-of-machine-learning-in-networking [Accessed: 12 February 2024].

Chen, J.I.-Z. & Lai, K.-L. (2021) Deep Convolution Neural Network model for credit-card fraud detection and alert, *June 2021*, 3(2), pp. 101–112. doi:10.36548/jaicn.2021.2.003.

Dzakiyullah, N. (2021) Semi-supervised classification on Credit Card Fraud Detection using autoencoders, *Journal of Applied Data Sciences*, 2(1), pp. 1–7. doi:10.47738/jads.v2i1.16.

Eliaçık, E. (2022) *Artificial Intelligence vs. human intelligence: Can a game-changing technology play the game?*, *Dataconomy*. Available at: https://dataconomy.com/2022/04/20/is-artificial-intelligence-better-than-human-intelligence/ [Accessed: 25 January 2024].

Elucidate (2023) *What the Danske Bank scandal can teach US about financial crime risk management in correspondent banking*, *Elucidate*. Available at: https://www.elucidate.co/blog/what-the-danske-bank-scandal-can-teach-us-about-financial-crime-risk-management-in-correspondent-banking#:~:text=The%20money%20laundering%20scheme%2C%20which,various%20accounts%20around%20the%20world. [Accessed: 12 February 2024].

Intellias (2022) *How to use machine learning in fraud detection*, *Intellias*. Available at: https://intellias.com/how-to-use-machine-learning-in-fraud-detection/#:~:text=Fraud%20detection%20using%20machine%20learning%20can%20solve%20all%20of%20these [Accessed: 04 February 2024].

Flavián, C., Pérez-Rueda, A., Belanche, D. & Casaló, L.V. (2021) Intention to use Analytical Artificial Intelligence (AI) in services – the effect of technology readiness and awareness, *Journal of Service Management*, 33(2), pp. 293–320. doi:10.1108/josm-10-2020-0378.

Fraud.com (2023) *The advantages of machine learning in fraud prevention*, *Fraud.com*. Available at: https://www.fraud.com/post/the-advantages-of-machine-learning-in-fraud-prevention [Accessed: 09 February 2024].

Forbes (N.D.) *Danske Bank | Company Overview & News*, *Forbes*. Available at: https://www.forbes.com/companies/danske-bank/ [Accessed: 7 February 2024].

Gee, J., & Button, M. (2019) The Financial Cost of Fraud 2019. *Technology Paper*. Portsmouth: University of Portsmouth, pp. 1–25. Available at: http://www.crowe.ie/wp-content/uploads/2019/08/The-Financial-Cost-of-Fraud-2019.pdf [Accessed: 27 January 2024].

GfG (2023) *Machine Learning for Anomaly Detection*, *GeeksforGeeks*. Available at: https://www.geeksforgeeks.org/machine-learning-for-anomaly-detection/ [Accessed: 8 February 2024].

Gottschalk, P. 2022. 6 Danske Bank in Denmark. The Convenience of Corporate Crime: Financial Motive – Organizational Opportunity – Executive Willingness. Berlin, Boston: De Gruyter, pp. 115-129. doi.org/10.1515/9783110766950-007

Khusheef, A.S., Shahbazi, M. & Hashemi, R. (2022) Investigation of long short-term memory networks for real-time process monitoring in fused deposition modeling, *Progress in Additive Manufacturing*, 8(5), pp. 977–995. doi:10.1007/s40964-022-00371-x.

Kureishy, A., Meley, C. & Mackenzie, B. (2018) *Achieving real business outcomes from Artificial Intelligence*, *O'Reilly Online Learning*. Available at: https://learning.oreilly.com/library/view/achieving-real-business/9781492038214/ [Accessed: 15 January 2024].

Kute, D.V., Pradhan, B., Shukla, N. & Alamri, A. (2021) Deep learning and explainable artificial intelligence techniques applied for detecting money laundering–A critical review, *IEEE Access, 9*, pp. 82300–82317. doi:10.1109/access.2021.3086230.

Lieberwitz, M. (2023) *Generative AI highlights the need for identity verification*, *insideBIGDATA*. Available at: https://insidebigdata.com/2023/10/27/generative-ai-highlights-the-need-for-identity-verification/ [Accessed: 10 February 2024].

LRT (2019) *Danske Bank ordered to close branch in Estonia, Will Keep Service Centre in Lithuania*, *lrt.lt*. Available at: https://www.lrt.lt/en/news-in-english/19/1119758/danske-bank-ordered-to-close-branch-in-estonia-will-keep-service-centre-in-lithuania [Accessed: 10 January 2024].

Manchev, N. (2021) *Credit card fraud detection using XGBoost, smote, and threshold moving*, *Domino Data Lab*. Available at: https://domino.ai/blog/credit-card-fraud-detection-using-xgboost-smote-and-threshold-moving [Accessed: 01 February 2024].

Milne, R. (2022) *Danske Bank to pay $2bn penalty for defrauding US banks*, *Financial Times*. Available at: https://www.ft.com/content/6a17f771-7c13-43f1-9d11-2e401db8e48f [Accessed: 14 February 2024].

Miroshnyk, O. (2022) *What is text analysis?*, *OneAI*. Available at: https://oneai.com/learn/text-analysis [Accessed: 09 February 2024].

Odmark, J. (2023) *What is Adaptive ML (online machine learning)?*, *Pandio*. Available at: https://pandio.com/what-is-adaptive-ml-online-machine-learning/#:~:text=Adaptive%20machine%20learning%20is%20a,and%20provides%20insights%20almost%20instantaneously. [Accessed: 10 February 2024].

Planque, T. de (2017) *Big Data: Computer vs. human brain: MS&E 238 blog*, *MSE 238 Blog Big data Computer vs Human Brain Comments*. Available at: https://mse238blog.stanford.edu/2017/07/teun/big-data-computer-vs-human-brain/ [Accessed: 04 February 2024].

Schwartzkopff, F. (2021) *Danske's customer flight exposes long-term cost of a scandal*, *Bloomberg.com*. Available at: https://www.bloomberg.com/news/articles/2021-07-25/danske-s-customer-flight-exposes-the-long-term-cost-of-a-scandal [Accessed: 10 February 2024].

Shetty, S.H., Shetty, S., Singh, Ch., & Rao, A. (2022) Supervised machine learning: Algorithms and applications, *Fundamentals and Methods of Machine and Deep Learning*, pp. 1–16. doi:10.1002/9781119821908.ch1.

Singh, G. (2024) *A gentle introduction to autoencoders for data science enthusiasts*, *Analytics Vidhya*. Available at: https://www.analyticsvidhya.com/blog/2021/06/autoencoders-a-gentle-introduction/ [Accessed: 16 February 2024].

Subrahmannian, S. (2023) *What is LSTM? introduction to long short term memory*, *Intellipaat*. Available at: https://intellipaat.com/blog/what-is-lstm/ [Accessed: 15 February 2024].

Torabi, H., Mirtaheri, S.L. & Greco, S. (2023) Practical autoencoder based anomaly detection by using vector reconstruction error, *Cybersecurity*, 6(1). doi:10.1186/s42400-022-00134-9.

U.S. Department of Justice (2022) *Danske Bank pleads guilty to fraud on U.S. banks in multi-billion dollar scheme to access the U.S. Financial System*, *Office of Public Affairs* | *Danske Bank Pleads Guilty to Fraud on U.S. Banks in Multi-Billion Dollar Scheme to Access the U.S. Financial System* | *United States Department of Justice*. Available at: https://www.justice.gov/opa/pr/danske-bank-pleads-guilty-fraud-us-banks-multi-billion-dollar-scheme-access-us-financial [Accessed: 07 February 2024].