International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

The Age of Financial Frauds and using Random Forest Machine Learning to Predict Fraudulent Transactions

Dilsher Singh

dilsher.singh10[at]gmail.com

Abstract: ***With the digitalization of financial processes and the rapid growth of the fintech industry, there are large volumes of data, sensitive personal information and monetary exchanges in billions, resulting in a rapid growth of cases of financial fraud. The involvement of machine learning in analyzing customer data, extraction of information and the detection of patterns leads to more efficient and accurate predictions and results, allowing businesses and analysts to make decisions and take security measures accordingly. This research study focuses on using the Random Forest Machine Learning algorithm to detect and predict fraudulent transactions using a real-world dataset. The paper discusses stepwise approaches to process the data, train the model and then use the model to make predictions. The accuracy of prediction was found to be 99.94% with this algorithm; key aspects about the accuracy and run time of the algorithm have also been highlighted. This paper reinforces how implementing machine learning models in real-world transactional platforms would act as a safety net for consumers and businesses to prevent cyber security breaches.***

Keywords: Machine learning, random forest, fraudulent transactions, cybersecurity

1. Introduction

Machine learning (ML) is a powerful tool within the artificial intelligence umbrella that has transformed the competence of modern computing systems. With the help of a wide set of algorithms, ML provides computers with the ability to "perform intelligent predictions" (Nichols, Chan and Baker, 2018) using datasets. Its use revolves around using training data to build analytical models and enhance problem-solving and decision making (Janiesch, Zschech and Heinrich, 2021).These models can be used in various settings, allowing businesses to assess trends, analyze consumer behavior and patterns in operations. It also has significant applications in the health, finance and retail industries. For instance, models within ML can create substantial impact in the finance sector and cybersecurity by detecting and predicting fraudulent transactions. This report focuses on such occurrences, thereby implementing supervised machine learning models, namely the Random Forest algorithm, in transactional fraud detection with the help of real world data.

Credit card usage has increased rapidly and has resulted in the "escalation of fraud" (Sulaiman, Schetinin and Sant, 2022). E-payments, digital financial platforms and online transactions have significant roles to play in the contemporary financial ecosystem, making transaction processes more efficient. According to an article issued in Forbes Advisor, 28% of all transactions in 2021 were made using credit cards and there were approximately 485 million credit card accounts in total during the same year (Pokora, 2022). Furthermore, a press release by the World Bank states that digital payment methods are used worldwide by two-thirds of adults (worldbank.org, 2022). There is no doubt that credit cards and digital methods of payment are efficient modes of transaction, however, they are linked to security problems such as fraud, since sensitive information can be stolen by scammers and identity thieves. In 2021, there were 1.7 million cases of identity theft and 390 million people were impacted by credit card fraud in the United States alone (Daly and Caporal, 2022). The benefit of machine learning in the prediction and detection of fraudulent transactions lies in its capability to be able to extract "insights from data in real-time" (Kanade, 2022) and allow informed decision making based on patterns. ML is growing to be more relevant in today's digital age due to the massive amounts of data being generated every minute (Nichols, Chan and Baker, 2018), (Vadapalli, 2021).

Machine learning has provided computing systems and software applications with the ability and accuracy to predict outcomes without having been programmed explicitly. With the use of algorithms, input data is used to produce fresh output values. ML uses advanced problem solving techniques with the help of algorithmic models to create "predictions, rules, answers, recommendations" (Janiesch, Zschech and Heinrich, 2021). Algorithmic structures and approaches within machine learning can be classified into four pillars: supervised, un-supervised, semi-supervised and reinforcement learning (Kanade, 2022).

1.1. Types of Machine Learning

1.1.1. Supervised machine learning: It uses labeled data, allowing the training to guide the algorithmic prediction, with the parameters having already been mapped. Its most significant goal is to map the input and output variables and establish patterns with one another. We can broadly categorize supervised machine learning into *classification* and *regression* models (Pant, 2019). While the classification problems focus on assigning the output variable to a category based on its features, the regression problems tend to establish a linear relationship between two variables usually for the purpose of the prediction of output variables (Kanade, 2022).

1.1.2. Unsupervised machine learning: It focuses on making sense of unlabeled datasets. The model uses algorithms to assess the data and predict the output without supervision or having been trained. The algorithm organizes the data on the basis of the patterns including the similarities and differences. This type of machine learning can be classified broadly into *clustering* and *association* methods. While the clustering model relies on understanding the similarities or differences to group data into clusters, the association model identifies relationships between the variables within an extensive dataset (Kanade, 2022).

1.1.3. Semi-supervised machine learning: In these models, an amalgamation of the above two types is there - algorithms are trained using a combination of labeled and unlabeled datasets (Kanade, 2022).

1.1.4. Reinforced learning: It is based on behavior and feedback of the algorithm. It is enhanced according to its experiences as a result of positive and negative reinforcement. A good action is 'rewarded' with a positive signal while a bad action is associated with a negative one. The algorithm aims to maximize the positive signals and rewards enabling it to make fewer mistakes (Heidenreich, 2018).

1.2. The Random Forest Algorithm

In this report, the focus is on supervised machine learning, namely, the Random Forest algorithm. This ensemble learning technique combines classifiers for solving problems. It consists of a series of tree classifiers or decision trees; since the outcome depends on the prediction of the decision trees, an increase in the number of trees gives more precise outcomes (Liu, Whang, Zhang, 2012), (Mbaabu, 2022). The Random Forest model focuses on improving the accuracy of the prediction of a dataset by finding the average of the dataset. A training dataset in a decision tree model is divided into branches that continue to segregate till the algorithm results in a leaf node. The algorithm starts out with a root node (see Fig. 1), then branches out according to the "best split" (Reis, Baron and Shahaf, 2018), into decision nodes that ultimately result in a root node.

In most cases, Random Forest uses the 'bagging' model where multiple data samples are taken into consideration in order to result in accurate predictions. Here, a random sample is selected from the dataset and each model requires independent training. There is variation in the outputs and accuracy depending on the number of decision trees. Within a decision tree, unlabeled objects are analyzed based on their feature values and conditions before they reach a "terminal code" (Reis, Baron and Shahaf, 2018). The class with the best probability is termed as the predicted class. To put it simply, outputs go through a ranking process and the one with the highest rank is regarded as final (Mbaabu, 2022). Industrial applications of the Random Forest algorithm can be seen in the finance, healthcare and e-commerce industries. Some significant instances where this model would apply include the diagnosis of patients by assessing past records, analyzing stock market behavior, understanding and predicting consumer behavior and most relevant to this report, detecting cases of financial fraud.

This paper focuses on using the Random Forest supervised machine learning algorithm to detect and predict fraudulent transactions with the help of a real-world dataset. The paper discusses the importance of cybersecurity in today's data-driven world, how machine learning can come to aid, the approach followed for training the model, the accuracy of the model and other aspects that need to be addressed for practical applications. Overall, the paper implicates how machine learning is transforming the world how we see it.

2. Methodology

The dataset used for the model was retrieved from *kaggle.com* (Kaggle.com. 2022.), a trusted online community offering collaboration opportunities and datasets on relevant topics. The Google Colab platform was used with the Python programming language for executing the codes.

2.1 Understanding the dataset

The dataset for "Fraudulent Transactions Prediction" was downloaded having 6,362,620 rows and 10 columns. This data was chosen considering the relevance of frauds in today's digital world as discussed earlier in the paper. However, only the first 1,048,576 rows were used for training the model, the reason being discussed later in the paper. The categories featured here include *step*, *type*, *amount*, *nameOrig*, *oldbalanceOrg*, *newbalanceOrg*, *nameDest*, *oldbalanceDest*, *newbalanceDest*, *isFraud*, *isFlaggedFraud*. The *step* column implies units of time, where each step represents one hour of time; the type category indicates the type of transaction: whether it's incoming cash, cash leaving an account, transfer, debit or payment. The *amount* category refers to the transactional amount while the *nameOrig* category represents the customer whom the transaction originated with. The *oldbalanceDest* and *newbalanceDest* categories refer to the balance before the transaction was made and the updated balance after the transaction, respectively. The *nameDest* category refers to the recipient of the transaction. The *oldbalanceDest* and *newbalanceDest* represent the recipients balance before the transaction was made and the recipients updated balance post the transaction, respectively. The *isFraud* category tells whether a particular transaction was fraudulent or not. The *isFlaggedFraud* category highlights illegal attempts to transfer large sums of money.

The process starts with importing various Python libraries. Here is a brief about what these libraries contribute in the code. The NumPy library is necessary to work with arrays, the pandas library along with seaborn is for data analysis and matplotlib for data visualization including plotting graphs. LabelEncoder and train_test_split were imported for encoding the categorical (non-numerical) values and for training the dataset respectively.

The dataset was loaded into the pandas dataframe. Figure 1.1. shows the first five rows and 10 columns of the dataset being used. The df.shape function was used to express the dimensions of the data and inform about the number of rows and columns in it.

```
[ ] # first 5 rows of the dataframe
    df.head()
```

	step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
0	1	PAYMENT	9839.64	C1231006815	170136.0	160296.36	M1979787155	0.0	0.0	0.0	0.0
1	1	PAYMENT	1864.28	C1666544295	21249.0	19384.72	M2044282225	0.0	0.0	0.0	0.0
2	1	TRANSFER	181.00	C1305486145	181.0	0.00	C553264065	0.0	0.0	1.0	0.0
3	1	CASH_OUT	181.00	C840083671	181.0	0.00	C38997010	21182.0	0.0	1.0	0.0
4	1	PAYMENT	11668.14	C2048537720	41554.0	29885.86	M1230701703	0.0	0.0	0.0	0.0

Figure 1.1: Output showing the first five rows of the data set

In order to analyze the dataset further, *df.info( )* function (see Fig. 1.2) was used to print information about the dataset. This information would be imperative to understanding the data - number of columns, labels, data type, range index and the amount of cells. This also helped in gaining an understanding of the *non-null count* within the dataframe and assessing any missing values. If any, the imputation method would be helpful, which involves guessing the values, or removing the data of missing values altogether. Most importantly, this step made it easier to filter the categorical features that would be significant to this study. The categorical features included: *Type*, *nameOrig* and *nameDest.* The next step focused on confirming whether there were any missing values, by feeding in the *df.isnull( ).sum( ) function* (see Fig. 1.3). No missing values were found in our dataframe.

```
[ ] # getting some informations about the dataset
    df.info()

    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 1048575 entries, 0 to 1048574
    Data columns (total 11 columns):
     #   Column          Non-Null Count    Dtype
    ---  ------          --------------    -----
     0   step            1048575 non-null  int64
     1   type            1048575 non-null  object
     2   amount          1048575 non-null  float64
     3   nameOrig        1048575 non-null  object
     4   oldbalanceOrg   1048575 non-null  float64
     5   newbalanceOrig  1048575 non-null  float64
     6   nameDest        1048575 non-null  object
     7   oldbalanceDest  1048575 non-null  float64
     8   newbalanceDest  1048575 non-null  float64
     9   isFraud         1048575 non-null  int64
     10  isFlaggedFraud  1048575 non-null  int64
    dtypes: float64(5), int64(3), object(3)
    memory usage: 88.0+ MB
```

Categorical features:

- Type
- nameOrig
- nameDest

Figure 1.2: Output showing information about the datatype of every column in the dataset

```
[ ] # checking for missing values
    df.isnull().sum()

    step              0
    type              0
    amount            0
    nameOrig          0
    oldbalanceOrg     0
    newbalanceOrig    0
    nameDest          0
    oldbalanceDest    0
    newbalanceDest    0
    isFraud           0
    isFlaggedFraud    0
    dtype: int64
```

Figure 1.3: Output showing the sum of missing values in every column.

To visualize the data, the *Plots* function within pandas was used. The *sns.set( )* Plot function (see Fig. 1.4) was implemented, which is part of the Seaborn toolkit. It was established that there are no major inconsistencies in the data and all columns are relevant in the dataset. This eradicated possibilities of removing any column or deleting any feature from the dataset.

```
#Plots of numerical features
sns.set() #used for plots

#Distribution of Item_Weight
plt.figure(figsize =(6,6))
sns.countplot(x='type', data= df)
plt.show()
```

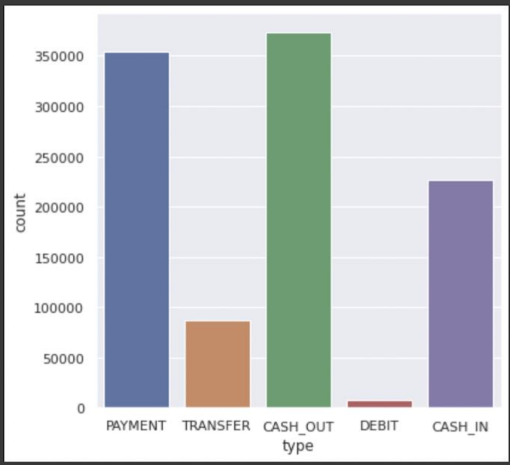

Figure 1.4: Output showing plot of Type column of the dataset

2.2. Data Processing

Standardization of data is a significant step in this research study as it gives us an opportunity to apply transformations to the dataset before it is fed into the algorithm. This step within the preprocessing workflow brings utility functions onboard by scaling the data to a unit variance. The focus shifted to adjusting vectors within the dataset and presenting it in a manner that would allow it to communicate with downstream estimators appropriately. The *LabelEncoder( )* function (see Fig. 1.5) was implemented in order to transform the categorical features into a numerical format to optimize it for algorithmic readability and allow it to communicate with the model. The *type*, *nameOrig* and *nameDest* categories were encoded (see Fig. 1.5). This was followed by replacing these encoded columns with the categorical columns in the original dataset. As shown in Fig. 1.6, this step successfully made the dataset free from categorical features.

```
[ ] from sklearn.preprocessing import LabelEncoder
    le = LabelEncoder()
    df['Type_Encoded'] = le.fit_transform(df['type'])
    df['nameOrig_Encoded'] = le.fit_transform(df['nameOrig'])
    df['nameDest_Encoded'] = le.fit_transform(df['nameDest'])
```

Figure 1.5: Encoding the categorical features of the dataset

```
[ ] df.info()

    <class 'pandas.core.frame.DataFrame'>
    RangeIndex: 1048575 entries, 0 to 1048574
    Data columns (total 11 columns):
     #   Column            Non-Null Count    Dtype
    ---  ------            --------------    -----
     0   step              1048575 non-null  int64
     1   amount            1048575 non-null  float64
     2   oldbalanceOrg     1048575 non-null  float64
     3   newbalanceOrig    1048575 non-null  float64
     4   oldbalanceDest    1048575 non-null  float64
     5   newbalanceDest    1048575 non-null  float64
     6   isFraud           1048575 non-null  int64
     7   isFlaggedFraud    1048575 non-null  int64
     8   Type_Encoded      1048575 non-null  int64
     9   nameOrig_Encoded  1048575 non-null  int64
     10  nameDest_Encoded  1048575 non-null  int64
    dtypes: float64(5), int64(6)
    memory usage: 88.0 MB
```

Figure 1.6: Output showing that all columns in the dataset have numerical values

2.3. Training the model

The train_test_split was used for training and testing the model. This function enabled the data arrays, represented by X (containing all features except *isFraud*) and Y(containing feature *isFraud*), to be split into two subsets, one for the training data and the other for test data. These subsets were then split at random, keeping 30% of the data for testing.

The SMOTE toolkit helps in balancing the data in case the proportion of data for both the categories is imbalanced.

The *Random Forest Classifier* algorithm (see Fig. 1.8), used for training the model, runs multiple trees using subsets from the training data at random. These decision trees provide the algorithm with votes and it combines them in order to calculate an accurate final result. The *n_estimators* function was used in order to specify the number of trees required, before the averaging process ensues. 1000 decision trees were chosen with the hope that it would lead to a more accurate prediction. In order to generate randomness, which would then lead to consistency and the reproduction of results in the prediction, the *random_state* parameter was used and its value specified as 2, in order to control the sets that would generate randomly. The run time for training the model was 1 hour 10 mins. The trained model was then used to predict Y values of the test dataset. The Y value being 0 or 1 would indicate whether a transaction is fraudulent or not. Metrics from the sklearn estimator were imported in order to quantify the functionality of the classification process and assess the accuracy of the prediction. The accuracy score was found to be **0.9993705753513493,** which is impressive (see Fig. 1.8)**.** This score is indicative of the fact that the model is trained at providing a very accurate result. This was possible since the Random Forest approach includes multiple estimates involving multiple data subsets.

```
[ ] from sklearn import metrics
    print("Accuracy:",metrics.accuracy_score(y_test, y_pred))

    Accuracy: 0.9993705753513493
```

Figure 1.8: Output showing accuracy of the model

3. Discussion

As noticed in this research study, the Random Forest model predicts fraudulent transactions with superior accuracy. Not only is it flexible, but it also allows analysts to assess the importance of variables within the model. Its high level of accuracy is a key reason as to why it is such a reliable tool; and its consistent use of averaging allows analysts to work with large volumes of data. While other models such as the Light Gradient Boosting Machine and Neural Networks are also popular amongst data scientists (Pykes, 2020), the Random Forest method, although time consuming, is one of the most versatile tools. It can deal with imbalanced data and handle numerical and categorical datasets with minimum pre-processing. For the current work, it was observed during running the code that the run time was very high. The runtime for training the model was 1 hour 10 minutes for only 1,048,576 rows of data. Since the random forest algorithm uses multiple sub-datasets, the run time is expected to be on the higher side. In real-world datasets, where systems deal with billions of rows of data, this algorithm may not prove to be an efficient way. In order to deal with overfitting and runtime challenges, one can use a smaller number of trees or lower the amount of estimators involved while analyzing metrics. By observing the top predictors and their significance to the model as it begins converging, one can make the decision to add more trees. While adding more trees would result in a more accurate prediction, using tuning parameters for specifying the minimum or maximum number of features during the step where the splitting of nodes takes place, would help in optimisation and prevent overfitting.

Despite various challenges, innovations knock the doors every hour and the relevance of machine learning in modern day industries is growing more prominent. 67% of companies use machine learning (Brown, 2021) including globally known companies like PayPal are also relying on ML to improve their fraud detection and risk management capabilities (Srivastava, 2022), thus growing safer from financial loss. Large economies are tackling financial frauds by setting up digital agencies that work in sync with financial institutions and law enforcement bodies. For example, the government of India is planning to establish a nodal agency called “Digital Intelligence Unit” (Government of India, 2021) in order to tackle digital frauds and investigate them.

The role of machine learning can be seen significantly in the fintech and cybersecurity sectors. Not only does ML play a prominent role in automating tasks, but it identifies and tackles attacks related to cyber security. It allows analysts to detect as well as classify anomalies and threats by identifying patterns (Deshmukh, 2020). It assists in the analysis of malware, phishing, making predictions and “clustering security events” (Linders, 2021). Predictive analysis models have been able to protect personal data and billions of critical transactions (Srivastava, 2022). Not only has machine learning reduced the number of instances where fraudulent transactions have been flagged inaccurately or where cases of transactional fraud are overlooked while also eliminating risk of human error, but it has also lead to the quicker collection of data, efficiency in handling large volumes of data and the effortless tackling of breaches in security.

4. Conclusion

Our world is in the prime of its digital age, with technology and artificial intelligence significantly impacting our day to day lives and revolutionizing businesses. The Random Forest model is a robust machine learning algorithm that is useful for generating solutions to problems related to classification. With the help of decision trees, randomization, categorical assessments, using multiple subsets and a ‘voting’ process as well as an in depth filtration and analysis of data frames, this model is a reliable method to receive accurate predictions. With the help of the Random Forest model, this study was able to generate accurate predictions as to whether a transaction is fraudulent or not. A real-world dataset was used to train the Random forest algorithm and predict accurate outcomes. This model is promising and has major scope in various industries, especially in solving issues related to cybercrime and financial fraud.

References

[1] Bin Sulaiman, R., Schetinin, V. and Sant, P., 2022. Review of Machine Learning Approach on Credit Card Fraud Detection. *Human-Centric Intelligent Systems*, 2(1-2), pp.55-68.

[2] Brown, S., 2021. *Machine learning, explained | MIT Sloan*. [online] MIT Sloan. Available at: <https://mitsloan.mit.edu/ideas-made-to-matter/machine-learning-explained> [Accessed 8 October 2022].

[3] Burns, E., 2022. *What Is Machine Learning and Why Is It Important?*. [online] SearchEnterpriseAI. Available at: <https://www.techtarget.com/searchenterpriseai/definition/machine-learning-ML> [Accessed 8 October 2022].

[4] Daly, L. and Caporal, J., 2022. *Identity Theft and Credit Card Fraud Statistics for 2022 |The Ascent*. [online] The Motley Fool. Available at: <https://www.fool.com/the-ascent/research/identity-theft-credit-card-fraud-statistics/> [Accessed 8 October 2022].

[5] Deshmukh, A., 2020. *Debunking the Myths. How Machine Learning (ML) Benefits Cyber Security*. [online] SecurityHQ. Available at: <https://www.securityhq.com/blog/debunking-the-myths-how-machine-learning-ml-benefits-cyber-security/> [Accessed 8 October 2022].

[6] Digitalindia.gov.in. 2021. *Digital Intelligence Unit Planned by Government to Help Deal With Fraudulent Transactions and Telecom Frauds | Digital India Programme | Ministry of Electronics & Information Technology(MeitY) Government of India*. [online] Available at: <https://digitalindia.gov.in/content/digital-intelligence-unit-planned-government-help-deal-fraudulent-transactions-and-telecom> [Accessed 8 October 2022].

[7] Education, I., 2022. *What is Random Forest?*. [online] Ibm.com. Available at: <https://www.ibm.com/cloud/learn/random-forest> [Accessed 8 October 2022].

[8] Heidenreich, H., 2018. *What are the types of machine learning?*. [online] Medium. Available at: <https://towardsdatascience.com/what-are-the-types-of-machine-learning-e2b9e5d1756f> [Accessed 8 October 2022].

[9] Janiesch, C., Zschech, P. and Heinrich, K., 2021. Machine learning and deep learning. *Electronic Markets*, 31(3), pp.685-695.

[10] Kaggle.com. 2022. *Fraudulent Transactions Prediction*. [online] Available at: <https://www.kaggle.com/datasets/vardhansiramdasu/fraudulent-transactions-prediction?datasetId=2014632&sortBy=dateRun&tab=profile> [Accessed 6 October 2022].

[11] Kanade, V., 2022. *What Is Machine Learning? Definition, Types, Applications, and Trends for 2022 | Spiceworks*. [online] Spiceworks. Available at: <https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/> [Accessed 8 October 2022].

[12] Kho, J., 2018. *Why Random Forest is My Favorite Machine Learning Model*. [online] Medium. Available at: <https://towardsdatascience.com/why-random-forest-is-my-favorite-machine-learning-model-b97651fa3706> [Accessed 8 October 2022].

[13] Linders, B., 2021. *What Machine Learning Can Do for Security*. [online] InfoQ. Available at: <https://www.infoq.com/news/2021/11/machine-learning-security/> [Accessed 8 October 2022].

[14] Liu, Y., Wang, Y. and Zhang, J., 2012. New Machine Learning Algorithm: Random Forest. *Information Computing and Applications*, pp.246-252.

[15] Mbaabu, O., 2020. *Introduction to Random Forest in Machine Learning*. [online] Engineering Education (EngEd) Program | Section. Available at: <https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/> [Accessed 8 October 2022].

[16] Melcher, K. and Silipo, R., 2019. *Fraud Detection Using Random Forest, Neural Autoencoder, and Isolation Forest Techniques*. [online] InfoQ. Available at: <https://www.infoq.com/articles/fraud-detection-random-forest/> [Accessed 8 October 2022].

[17] Muncaster, P., 2018. [online] Available at: <https://www.infosecurity-magazine.com/news/global-fraud-hits-32-trillion/> [Accessed 8 October 2022].

[18] Nichols, J., Chan, H. and Baker, M., 2018. Machine learning: applications of artificial intelligence to imaging and diagnosis. *National Library of Medicine*, 11.

[19] Pant, A., 2019. Introduction to Machine Learning for Beginners. *Towards Data Science*,.

[20] Pokora, B., 2022. *9 Interesting Credit Card Statistics*. [online] Forbes Advisor. Available at: <https://www.forbes.com/advisor/credit-cards/credit-card-statistics/> [Accessed 8 October 2022].

[21] Pykes, K., 2020. *Using Machine Learning To Detect Fraud*. [online] Medium. Available at: <https://towardsdatascience.com/using-machine-learning-to-detect-fraud-f204910389cf> [Accessed 8 October 2022].

[22] Reis, I., Baron, D. and Shahaf, S., 2018. Probabilistic Random Forest: A Machine Learning Algorithm for Noisy Data Sets. *The Astronomical Journal*, 157(1).

[23] Schonlau, M. and Zou, R., 2020. The random forest algorithm for statistical learning. *The Stata Journal: Promoting communications on statistics and Stata*, 20(1), pp.3-29.

[24] Srivastava, S., 2022. *How Machine Learning Helps in Financial Fraud Detection?*. [online] Appinventiv. Available at: <https://appinventiv.com/blog/role-of-machine-learning-in-financial-fraud-detection/> [Accessed 8 October 2022].

[25] Vadapalli, P., 2021. *Random Forest Classifier: Overview, How Does it Work, Pros & Cons | upGrad blog*. [online] upGrad blog. Available at:

<https://www.upgrad.com/blog/random-forest-classifier/> [Accessed 8 October 2022].

[26] worldbank.org. 2022. *COVID-19 Drives Global Surge in use of Digital Payments*. [online] Available at: <https://www.worldbank.org/en/news/press-release/2022/06/29/covid-19-drives-global-surge-in-use-of-digital-payments> [Accessed 8 October 2022].

Author Profile

Dilsher Singh is an Artist at Sunovatech. Having worked in the UI/UX and immersive reality industries, his career has closely been linked with upcoming technologies and creative digital solutions. Having been part of the digital revolution, Dilsher discovered a passion for artificial intelligence and machine learning. The scope of artificial intelligence in the modern cyber world is very promising as a result of industries relying heavily on this new technological paradigm. He has worked closely with companies such as Coca-Cola, Synapse India etc. He aims to dive deeper into the domain of artificial intelligence and significantly contribute to the world through innovations.