HIGH SCHOOL EDITION
Journal of Student Research
Volume 11 Issue 3 (2022)

How Efficient is Machine Learning in Detecting Financial Fraud using Mobile Transaction Metadata?

Vedant Shah[1]

[1]Cathedral & John Connon School, India

ABSTRACT

Global data has shown that each day the number of fraudulent transactions is exponentially increasing. As digital fraud continues to increase, machine learning and AI are being used to curb this increase. This paper will focus on the development and analysis of a machine learning model used to detect financial fraud in mobile money transactions. The aim of this paper is to discuss how effective a machine learning model can be in detecting fraud. This algorithm makes use of the Naïve Bayes model to detect fraud. To prove this, the results have been provided by using a confusion matrix, precision and recall which indicate the effectiveness of the model. The accuracy of the developed model is 0.996 with a precision and recall of both above 0.99. This work explores and develops a solution to one of the biggest breaches in digital security.

Introduction

The use of technology is exponentially increasing every day since the world has moved online. Due to this, a massive spark in digital financial fraud has been seen globally. Financial transactions take place online via banking cards, internet banking and mobile banking to name a few[1]. The method of making payments online proves to be extremely advantageous due to its convenience, immediate transfers, and year-round accessibility[2]. However, as the use of online payments increases, so does the number of fraudulent transactions.[3] Every year, the European Union Cybersecurity Agency reports losing billions of pounds, usually indicating around 2 million fraudulent digital transactions. According to research by TransUnion, digital fraud attempts between the end of 2020 and start of 2021 rose by 149%.[4] Furthermore, due to the pandemic, more people are using online and mobile banking to make any payments. Therefore, due to the global extent of digital payments and simultaneously, the rise in online fraud, investigating and developing methods to detect financial fraud in digital payments is extremely significant.

In this paper, a transaction will be considered fraudulent if the amount of money leaving the client's account is not equal to the amount of money being added to the recipient's account. There are various ways in which a fraudulent transaction can occur, such as when there is a facility takeover.[5] A facility takeover is when a bank account or credit card is accessed by a fraudster leading to the misuse of the client's funds. Wire transfer fraud is another way this occurs. A wire transfer fraud is where corrupt insiders may use forged documents to request a client's funds be transferred to another account instead of the intended recipients, and many more. [6] Initially, fraudulent transactions were detected by tip lines where anonymous users would contact the organization regarding any suspicion in their transactions. According to the Association of Certified Fraud Examiners (ACFE) there was a 50% decrease in fraudulent transactions at organizations with tip lines.[7] However, such methods usually worked with larger organizations and companies. Today, individuals who make transactions between themselves are also experiencing financial fraud as explained earlier. Therefore, new, and improved methods have been developed to detect fraudulent transactions. Data analytics is being used worldwide to examine transactions. Models and algorithms are being developed in machine learning, to test trends in fraudulent transactions using big data. However, these methods are not completely accurate and still result in many false positives and false negatives. Hence, this work will discuss the efficiency of Machine Learning in detecting financial fraud using metadata.

This paper will test and examine a 'paysim synthetic dataset' of mobile money transactions[8]. In this work, a model had been developed based on the Naïve Bayes Theorem to detect whether a transaction is fraudulent or not. The model will be tested using the paysim dataset. The accuracy of the model will be explained via a confusion matrix and graph in the results.

To achieve the stated research goals, this work is organized as follows: Section 2 will have a literature review explaining past papers based on topics similar to this work and will explain the difference and significance of this paper as compared to the other reports. Section 3 will discuss and explain the model employed in the algorithm. Section 4 will speak about the results and evaluate the algorithm. Section 5 will discuss the limitations of the model and algorithm and section 6 will be an overall conclusion of the paper.

Literature Review

The global fraud detection market is growing exponentially due to the increase in financial fraud. Originally valued at $29.8 Billion, according to VynZ Research, the market is predicted to be worth $92.3 Billion by 2027[9]. Two years ago, the Consumer Sentinel work received over 3.2 million records of online fraud, and these numbers have been rising ever since[10]. Numerous scholars and financial analysts have written papers based on this issue, similar to the topic being investigated in this work. The most commonly used method in predicting digital fraud is by using a data set to train a machine learning model. Based on this approach, there are papers which speak about the prediction aspect of financial fraud and some which speak about the generation of accurate datasets.

Alonso(2016)[11] discusses the lack of available datasets to test financial fraud models and proposes a solution known as the PaySim simulator. This simulator simulates transactions based on an original dataset to create a new, mostly accurate, dataset that may be used to train machine models. In this research paper, the accuracy, testing methods and results of the PaySim simulator are expalined. It makes use of a synthetic PaySim simulator to test the algorithm in detecting financial fraud.

Chyan-long Jan (2018)[12] establishes a model to detect financial fraud for the sustainable development of various enterprises and financial markets. The data set is employed is made up of companies listed on the Taiwanese Stock Exchange. Chyan-long Janc(2018), similar to this work, uses artificial intelligence techniques to predict fraud. However, it employs and evaluates more than 1 method such as an artificial neural network (ANN), decision trees etc. and compares the accuracy of them. In this paper, only one machine learning technique based on the Naïve Bayes Model will be employed and the effect of different attributes on the resulting accuracy will he compared in the evaluation.

Department of Banking and Financial Services at Kharkhiv National University of Economics and the Department of Finance and Banking at Pryazovskyi State Technical University (2020)[13] uses Big Data analysis algorithms and automated machine learning to develop effective models for detecting fraud in digital payment systems. Similar to this work, the paper displays its results using a confusion matrix, precision and recall. However, this paper focuses on various experiments whereas this work focus on a detailed evaluation and accuracy of one ML model test.

A paper in the Frontiers of Business Research in China (2020)[14] evaluates different machine models, however, does not establish its own model and algorithm. This paper is more of a literature review and evaluative report and uses true/false positive/negatives in its evaluations. Similarly, this paper will make use of TP, TN, FP, FN in the results section to determine the accuracy of the model.

Overall, this paper contains similarities with various other scholarly articles, such as but not limited to the ones mentioned above. However, it is different in the aspect where it is one of few papers using probability (Bayes Theorem) to evaluate just one machine learning model and determine its accuracy in detecting financial fraud.

Method and Data

This work has employed the Naïve Bayes Model to detect mobile transaction fraud. This paper will test and examine a 'paysim synthetic dataset of mobile money transactions. The code to the algorithm has been written using Jupyter, a python coding software.

Naïve Bayes Model

The Naïve Bayes Model is a simple classification technique based on Bayes' Theorem. Bayes Theorem is a mathematical formula to determine conditional probability[15]. Conditional probability is the probability of an event occurring which has a relationship to at least one or more other events, or in simpler terms, is the probability of an event occurring based on an event that has already occurred[16]. The term Naïve is used here because all features being tested are considered to be independent of one another and equal. This will be explained later on with the dataset.

Naive Bayes Formula

$$P(A|B)=\frac{P(B|A)P(A)}{P(B)}$$

Equation 1. The Naïve Bayes Formula

Equation 1 results in the conditional probability that event A takes place, given that event B is true.

- P(A)→ The probability that event A takes place before considering any another evidence. This is also known as the prior probability of A.
- P(B)→ This is the prior probability of the predictor or event B.
- P(B|A)→ This is the conditional probability of event B given that event A was true.
- P(A|B)→ This is the conditional probability of event A given that event B is true. It also known as posterior probability.

Consider the following made up real-world example to understand the formula:

Table 1. Real world example

	False transactions	True transactions	Total
Mobile wallet payments	4	12	16
Internet banking	6	13	19
Total	10	25	35

Table 1 shows the number of false and true transactions for two different payment methods: mobile wallet payments and internet banking.

Finding the likeliness of a transaction being true if it is made via internet banking, using the Naïve Bayes Formula:

- P(A)/P(True transactions)→ 25/35= 0.713
- P(B)/P(Internet banking)→19/35= 0.543
- P(B|A)/P(Internet banking | True Transactions)→ 13/25= 0.52

Therefore using the Naïve Bayes Formula:

$$P(\text{True Transactions} \mid \text{Internet Banking}) = \frac{P(\text{Internet banking} \mid \text{True Transactions}) * P(\text{True transactions})}{P(\text{Internet Banking})}$$

$$P(\text{Internet Banking} \mid \text{True Transactions}) = \frac{0.52 * 0.713}{0.543} = 0.682$$

Hence, the probability if a transaction made via internet banking will be true is 0.682, indicating that is likely for a transaction to be true if it is made via internet banking, in the case of this instructional example.
The Naïve Bayes Classifier has various variants such as Gaussian Naïve Bayes, Multinomial Naïve Bayes and Bernoulli Naïve Bayes. The *Gaussian Naïve Bayes* has been used for this paper. Gaussian Naïve Bayes can calculate continuous values, such as the data in our dataset, and assumes that every feature follows a normal distribution, or Gaussian distribution[17].

The dataset

A Paysim Synthetic dataset of mobile money transactions has been employed in this work. It shows continuous real time data. The dataset consists of steps in which one step is equal to one hour of real time simulation. The model will be tested and refined using this data. This dataset consists of various attributes or data columns:

step→ this attribute shows the real time simulation. It consists of '1' in each row, representing an hour of simulation.
type→ this attribute consists of the various types of mobile money transactions such as CASH_IN, CASH_OUT, DEBIT, PAYMENT and TRANSFER.
amount→ this attribute consists of the amount of money being transferred or used in the transaction in terms of the local currency.
nameOrig→ this attribute consists of the customer's account number.
oldbalanceOrg→ this attribute consists of the amount of money in the customer's account before the transaction.
newbalanceOrig→ this attribute consists of the amount of money in the customer's account after the transaction.
nameDest→ this attribute consists of recipient's account number.
oldbalanceDest→ this attribute consists of the amount of money in the recipient's account before the transaction.
newbalanceDest→ this attribute consists of the amount of money in the recipient's account after the transaction.
isFraud→ this attribute identifies if a transaction is fraudulent (1) or not (0).
isFlaggedFraud→ this attribute flags any transaction that tries to transfer over 200.00 in one transaction as fraudulent.

This dataset consists of 6362620 rows of data, each row consisting of one transaction, and 11 columns, each column representing a different attribute. Therefore, the total size of the dataset is 6362620 x 11.

Table 2. The dataset layout

step	type	amount	…	isFraud	isFlaggedFraud
1	PAYMENT	9839.64	…	0	0
1	PAYMENT	1864.28	…	0	0
1	TRANSFER	181.00	…	1	0

Table 2 is a layout of the first three rows of the dataset. As explained before, the attributes can be seen as the column headings and each row represents a different transaction. In this dataset, only the third row is classified as a fraudulent transaction due to the Boolean digit '1' indicating that 'isFraud' is true.

The Machine Learning Algorithm

Python is a programming language which is used to automate tasks and can carry out data analysis[18]. Python has been to code the machine learning model in this paper.

The algorithm works in the following manner:

1. The Paysim Synthetic Database of Mobile Money Transactions is imported onto the coding software.
2. The attributes are split into the independent and dependent variables. All attributes are taken as independent variables except 'isFraud' which is taken as a dependent variable. This is because the other attributes determine whether a transaction is fraudulent, where as the result in the 'isFraud' column is dependent on the values of the other variables.
3. The data is then split into a train set and a test set. The Naïve Bayes Model code is then trained and cleaned using the train set and eventually the model is tested using the test set (20% of the whole dataset) which gives the final results.

Results and Evaluation

The model was tested using each attribute of the dataset acting as an independent variable in the calculation. This section will explain the results and the effect different variables had on the results.

The Confusion Matrix

		True Values	
		Positive	*Negative*
Predicted Values	Positive	*TP*	*FP*
	Negative	*FN*	*TN*

Figure 1. The Confusion Matrix

Figure 1 represents an example of a confusion matrix. It shows how the results of an algorithm or model compare to the actual dataset it was tested with[19]. The matrix labels within the confusion matric in figure 2 have the following meaning in this paper:

TP (True Positive)→ This means that the result of a transaction in the dataset was fair and our model also identified it as a fair transaction.
FP (False Positive)→ This means that the result of a transaction in the dataset was fraudulent, but our model identified it as fair.
FN (False Negative)→ This means that the result of a transaction in the dataset was fair, but our model identified it as fraudulent.
TN (True Negative)→ This means that the result of a transaction in the dataset was fraudulent, and our model also identified it as a fraudulent.

Precision and Recall

Precision is the fraction of true positive's upon the total number of predicted positive's.[20] Therefore, in our dataset precision is the total number of transactions our model correctly identified as fair, upon the total number of transactions that were predicted to be fair.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Recall is the fraction of true positive's upon the total number of actual positive's.[21] Therefore, in our dataset, recall is the total number of transactions our model correctly identified as fair, upon the total number of transactions that were actually fair.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Final results

The model when using all attributes in its calculation, had the following outcome:

Accuracy: 0.996133 or 99.613%

True Positive: 1267408 transactions
False Positive: 3475 transactions
False Negative: 1444 transactions
True Negative: 197 transactions

Precision: $\frac{1267408}{1267408 + 3475} = 0.9973 = 99.7\%$

Recall: $\frac{1267408}{1267408 + 1444} = 0.9989 = 99.9\%$

Confusion Matrix

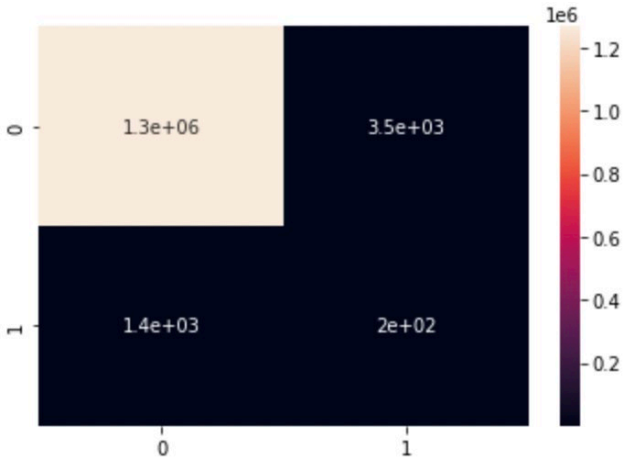

Figure 2. The resulting confusion matrix

From these results we understand that our model is highly accurate and will be able to tell if a transaction is fraudulent 99.6% of the time.

Evaluation

To better evaluate the model, a test was conducted on the model by calculating its prediction accuracy by removing some attributes. This is similar to conducting an AI ablation study.

CASE 1 (without 'type')

Accuracy: 0.99618 or 99.618%

		True Values	
		Positive	Negative
Predicted Values	Positive	1267469	3414
	Negative	1448	193

Figure 3. Confusion matrix without 'type'

Compared to the actual model, accuracy increased by 0.005%.

CASE 2 (without 'amount')

Accuracy: 0.99793 or 99.793%

		True Values	
		Positive	Negative
Predicted Values	Positive	1269887	996
	Negative	1636	5

Figure 4. Confusion matrix without 'amount'

Compared to the actual model, accuracy increased by 0.180%.

CASE 3 (without 'oldbalanceOrg' and 'newbalanceOrig)

Accuracy: 0.996132 or 99.6132%

		True Values	
		Positive	*Negative*
Predicted Values	Positive	1267412	3471
	Negative	1450	191

Figure 5. Confusion matrix without 'oldbalanceOrg' and 'newbalanceOrig'

Compared to the actual model, accuracy decreased by 0.0001%.

CASE 4 (without 'oldbalanceDest' and 'newbalanceODest)

Accuracy: 0.99671 or 99.671%

		True Values	
		Positive	*Negative*
Predicted Values	Positive	1268141	2742
	Negative	1450	191

Figure 6. Confusion matrix without 'oldbalanceDest' and 'newbalanceDest'

Compared to the actual model, accuracy increased by 0.058%.

CASE 5 (without 'isFlaggedFraud')

Accuracy: 0.99169 or 99.169%

		True Values	
		Positive	*Negative*
Predicted Values	Positive	1261662	9221
	Negative	1343	298

Figure 7. Confusion matrix without 'isFlaggedFraud'

Compared to the actual model, accuracy decreased by 0.444%.

CASE 6 (without 'oldbalanceOrg', 'newbalanceOrig', 'oldbalanceDest' and 'newbalanceDest')

Accuracy: 0.99669 or 99.669%

		True Values	
		Positive	*Negative*
Predicted Values	Positive	1268129	2754
	Negative	1449	192

Figure 8. Confusion matrix without 'oldbalanceOrg', 'newbalanceOrig', 'oldbalanceDest' and 'newbalanceDest'
Compared to the actual model, accuracy increased by 0.056%.

TABLE OF ACCURACY

Table 3. Comparison of results

CASE NUMBER	ACCURACY
Original Matrix	99.613%
CASE 1	99.618%
CASE 2	99.793%
CASE 3	99.613%
CASE 4	99.671%
CASE 5	99.169%
CASE 6	99.669%

The above table lists the different cases and their accuracies as compared to the original result where all attributes were used. This evaluation has shown that 'amount' plays the biggest role in leading to false results as when the amount was removed, the accuracy of the model increased the most. It also shows us that 'isFlaggedFraud' is the most necessary for accurate results as without this variable, there was a 0.444% decrease in accuracy.

Limitations

Machine learning algorithms are always susceptible to errors. This could be due to manual error, loopholes in formulas employed and even issues in data used. During this paper, there have been various limitations that might have tampered with the accuracy of the algorithm.

The dataset used, as stated before, consisted of over 6 million rows of data. This made it impossible to go through and check each row for any errors. There are chances some transactions might have been duplicated in the dataset, or that there might have been outliers which affected the calculation by the Naïve Bayes Formula.

As explained earlier, the Naïve Bayes classifier assumes all variables to be independent and equal in its calculations. However, this is never true in the real world and hence its calculations might not be able to exactly apply to the real world. Furthermore, wherever there is an empty cell in the dataset, it will assign the value zero which could

be different from the value intended to be in the dataset. However, while coding the algorithm, the dataset was cleaned which helped remove as many of these empty sets as possible. Furthermore, the calculations only lead to an estimation/prediction, hence it is not always accurate in the real world. It is extremely difficult to understand WHY each variable causes a fluctuation in the accuracy which makes it harder to understand why a transaction is fraudulent or not.

Conclusion

The basis of this paper was to discuss the efficiency of Machine Learning in detecting financial fraud using mobile transaction metadata. To answer this question, a dataset of mobile transactions was used to test a machine learning algorithm which was coded on Python. The algorithm used the Naive Bayes model to detected financial fraud.

Based on the results, it is fair to conclude that machine learning is an extremely efficient method to detect financial transactions due its success rate of over 99.6%. Furthermore, even though the algorithm does consist of few limitations, it uses real world, real time data and hence it's calculations and accuracy may be applied to the real world. The results of the confusion matrix, precision and recall all show how efficient the algorithm is in detecting fraud. Furthermore, machine learning proves to be advantageous as it carry out its operation in a very short period of time to calculate millions of rows of data as it did for this dataset. This paper also showed the effect of different attributes on the accuracy of the algorithm which helps understand issues in which variables contribute the most to making a transaction fraudulent.

Digital fraud is a rising issue and will continue to increase with the development of technology. The first step to alleviating this issue is to learn and understand how to detect it. Detection is the first step to prevention. By the development of efficient machine learning algorithms it is possible to understand when a transaction will be fraudulent. Cyber-crime, such as financial fraud, is more than just an issue that is affecting you or me, but is a power as big as one that can be used in warfare. It needs to be combatted and this paper has introduced one of many models that need to be further developed as the first step to stopping something bigger than financial fraud.

Acknowledgements

I would like to thank Vanessa Klotzman for guiding me through this project and providing me with the required data and assistance while training the machine learning model to analyse the dataset.

References

1. "Digital Payment - Different Modes & Benefits of Digital Payment." *Bankbazaar*, www.bankbazaar.com/ifsc/digital-payment.html.
2. DepositAccounts, et al. "Banking 101: 5 Benefits of Online Banking." *Deposit Accounts*, 22 Apr. 2019, www.depositaccounts.com/blog/online-banking.html.
3. *Financial Fraud in the Digital Space | Enisa - Europa*. www.enisa.europa.eu/publications/enisa-position-papers-and-opinions/financial-fraud-in-the-digital-space.
4. Smith, Tyler, et al. "Increase in Digital Banking Facilitating Equal Rise in Financial Fraud Attempts." *The Fintech Times*, 24 June 2021, https://thefintechtimes.com/increase-in-digital-banking-facilitating-equal-rise-in-financial-fraud-attempts/.
5. "Different Types of Financial Fraud Explained." *Financial Fraud Explained | Identity Theft and Fraud | Equifax UK*, www.equifax.co.uk/resources/identity-protection/different-types-of-financial-fraud-explained.html.

6. Jones, Marlene. "Fraudulent Wire Transfers: Examples of Fake Emails to Watch For." *Tetra Defense*, Marlene Jones Http://Www.tetradefense.com/Wp-Content/Uploads/2022/02/AW-TD_Logo_Landscape_Reverse-Blue-Copy-300x109.Png, 14 Nov. 2021, www.tetradefense.com/incident-response-services/fraudulent-wire-transfers-how-they-happen-and-why-you-may-be-at-risk/#:~:text=Wire%20transfer%20fraud%20has%20grown,to%20another%20person's%20bank%20account.
7. "5 Methods of Detecting Fraud in Organizations." *Kaufman Rossin*, https://kaufmanrossin.com/blog/5-methods-of-detecting-fraud-in-organizations/.
8. Arjunjoshua. "Predicting Fraud in Financial Payment Services." *Kaggle*, Kaggle, 21 Jan. 2018, www.kaggle.com/arjunjoshua/predicting-fraud-in-financial-payment-services/notebook.
9. Quinn, Megan. "Detecting Financial Fraud with Machine Learning." *Detecting Financial Fraud with Machine Learning*, www.bluegranite.com/blog/detecting-financial-fraud-with-machine-learning.
10. Techlabs, Maruti. "How Machine Learning Is Enhancing Fraud Detection." *Medium*, Medium, 16 Sept. 2020, https://marutitech.medium.com/how-machine-learning-is-enhancing-fraud-detection-694f3a2237f.
11. *PAYSIM: A Financial Mobile Money Simulator for Fraud Detection*. www.researchgate.net/publication/313138956_PAYSIM_A_FINANCIAL_MOBILE_MONEY_SIMULATOR_FOR_FRAUD_DETECTION.
12. Jan, Chyan-long. "An Effective Financial Statements Fraud Detection Model for the Sustainable Development of Financial Markets: Evidence from Taiwan." *MDPI*, Multidisciplinary Digital Publishing Institute, 14 Feb. 2018, www.mdpi.com/2071-1050/10/2/513.
13. Kolodiziev, Oleh, et al. "Automatic Machine Learning Algorithms for Fraud Detection in Digital Payment Systems." *SSRN*, 14 Jan. 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3730001.
14. Huang, Jian, et al. "Deep Learning in Finance and Banking: A Literature Review and Classification - Frontiers of Business Research in China." *SpringerOpen*, Springer Singapore, 8 June 2020, https://fbr.springeropen.com/articles/10.1186/s11782-020-00082-6.
15. "Bayes Theorem - Statement, Formula, Derivation, Examples." *Cuemath*, www.cuemath.com/data/bayes-theorem/.
16. Barone, Adam. "Conditional Probability." *Investopedia*, Investopedia, 8 Feb. 2022, www.investopedia.com/terms/c/conditional_probability.asp#:~:text=Conditional%20probability%20is%20defined%20as,succeeding%2C%20or%20conditional%2C%20event.
17. Majumder, Prateek. "Gaussian Naive Bayes." *OpenGenus IQ: Computing Expertise & Legacy*, OpenGenus IQ: Computing Expertise & Legacy, 23 Feb. 2020, https://iq.opengenus.org/gaussian-naïvebayes/#:~:text=Gaussian%20Naive%20Bayes%20is%20a,distribution%20and%20supports%20continuous%20data.&text=Naive%20Bayes%20are%20a%20group,technique%2C%20but%20has%20high%20functionality.
18. Coursera. "What Is Python Used for? A Beginner's Guide." *Coursera*, 22 Sept. 2021, www.coursera.org/articles/what-is-python-used-for-a-beginners-guide-to-using-python.
19. Markham, Kevin. "Simple Guide to Confusion Matrix Terminology." *Data School*, Data School, 3 Feb. 2020, www.dataschool.io/simple-guide-to-confusion-matrix-terminology/.
20. "Idiot's Guide to Precision, Recall, and Confusion Matrix." *KDnuggets*, www.kdnuggets.com/2020/01/guide-precision-recall-confusion-matrix.html.
21. Jayaswal, Vaibhav. "Performance Metrics: Confusion Matrix, Precision, Recall, and F1 Score." *Medium*, Towards Data Science, 15 Sept. 2020, https://towardsdatascience.com/performance-metrics-confusion-matrix-precision-recall-and-f1-score-a8fe076a2262