Team 5

Fraud Detection

Design Document

Table of Contents

[Introduction 2](#_Toc517345903)

[How Remitly Works 2](#_Toc517345904)

[Fraud types and algorithmic fraud detection logic 7](#_Toc517345905)

[Money Laundering 8](#_Toc517345906)

[Impersonation 8](#_Toc517345907)

[Terror Financing 9](#_Toc517345908)

[Elder Financial Abuse 9](#_Toc517345909)

[Machine Learning Algorithms 10](#_Toc517345910)

[Artificial Neural Network Back-Propagation Algorithm 12](#_Toc517345911)

[Self-organizing Mapping Neural Network 14](#_Toc517345912)

[Adaptive Neuro Fuzzy Approach 16](#_Toc517345913)

[System Design and Conclusion 17](#_Toc517345914)

[References 19](#_Toc517345915)

# Introduction

Wal-Mart, world’s largest company by revenue, has partnered with Remitly, Inc. to allow people to make person to person money transfers from the U.S., Canada, Australia and the UK to 10 developing countries through Remitly mobile app and Wal-Mart physical stores. Many of the recipients do not have bank accounts, but they can bring their mobile phone to a Wal-Mart and receive cash. Remitly transfers about $4 billion per year and has received $200 million in investor funding. The issue of fraud has been haunting the partnership between Wal-Mart and Remitly, especially the remittance payment scams. This report will look into the mechanism of Remitly’s wire transferring plus Wal-Mart’s cash pick-up process to investigate potential risks of fraud and provide solutions to alleviate such risks.

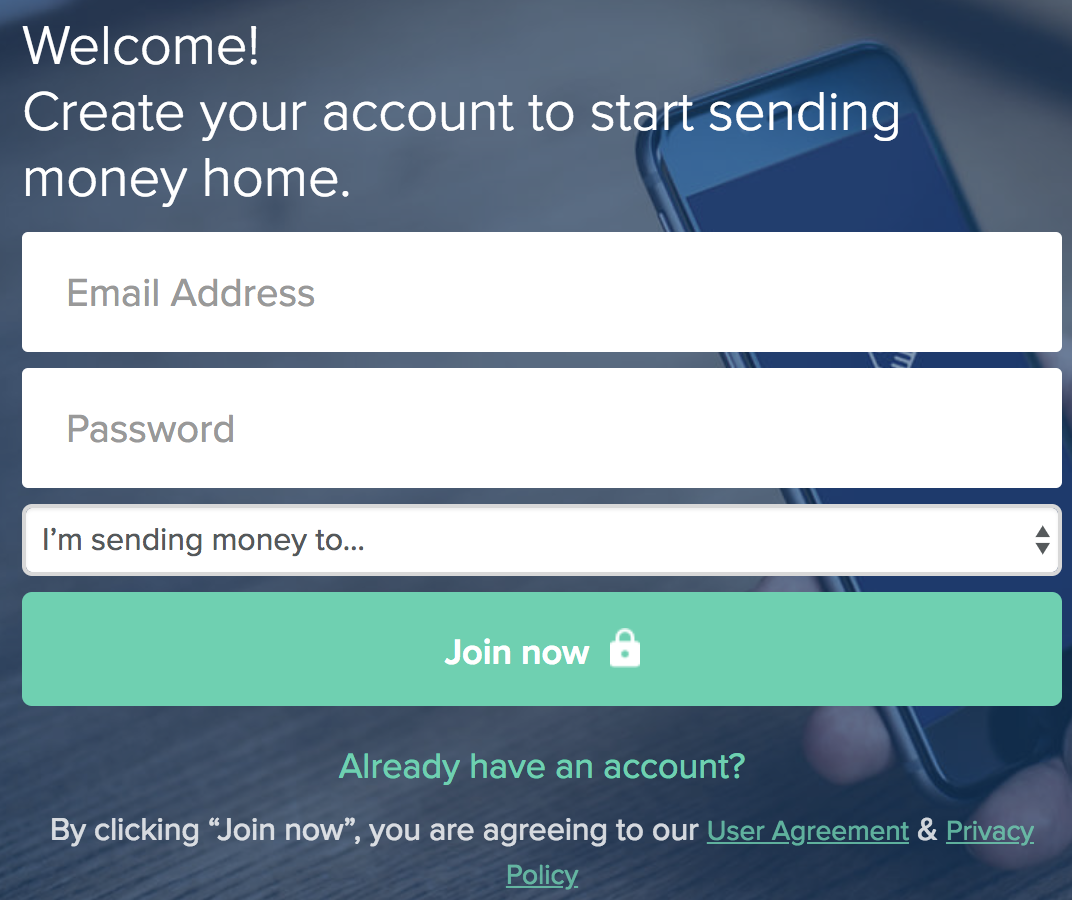
# How Remitly Works

Remitly is a payments company that leverages digital channels and mobile devices to send funds internationally. Remitly products are fast, inexpensive, and convenient for cross-border money transfers. Since fraud might exist from both the sender side and the recipient side, we walk through the steps of sending money and picking up cash in order to fully understand the whole process and to identify potential pitfalls.

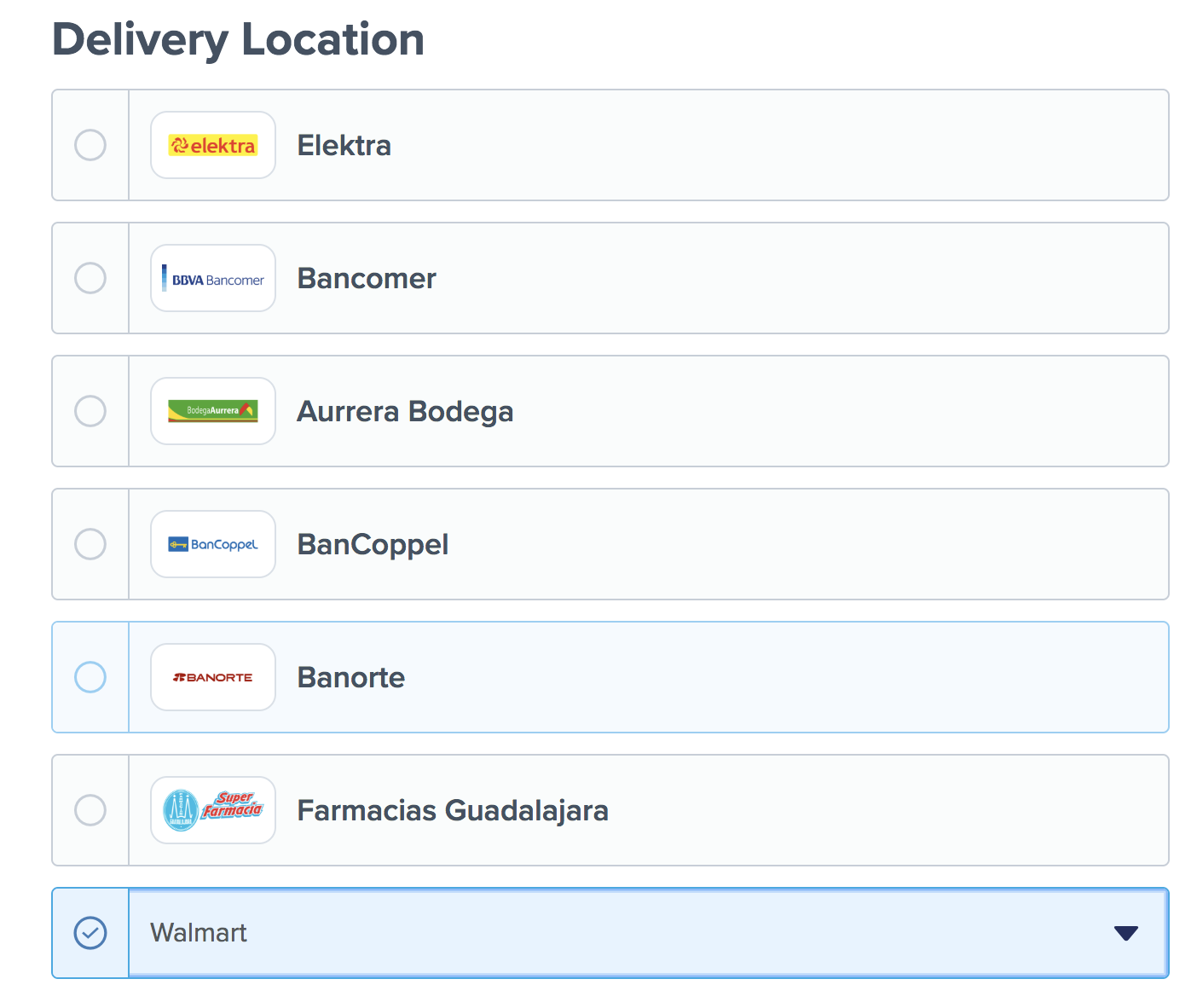
**Sending Money**

Customers take the following steps to make money transfers:

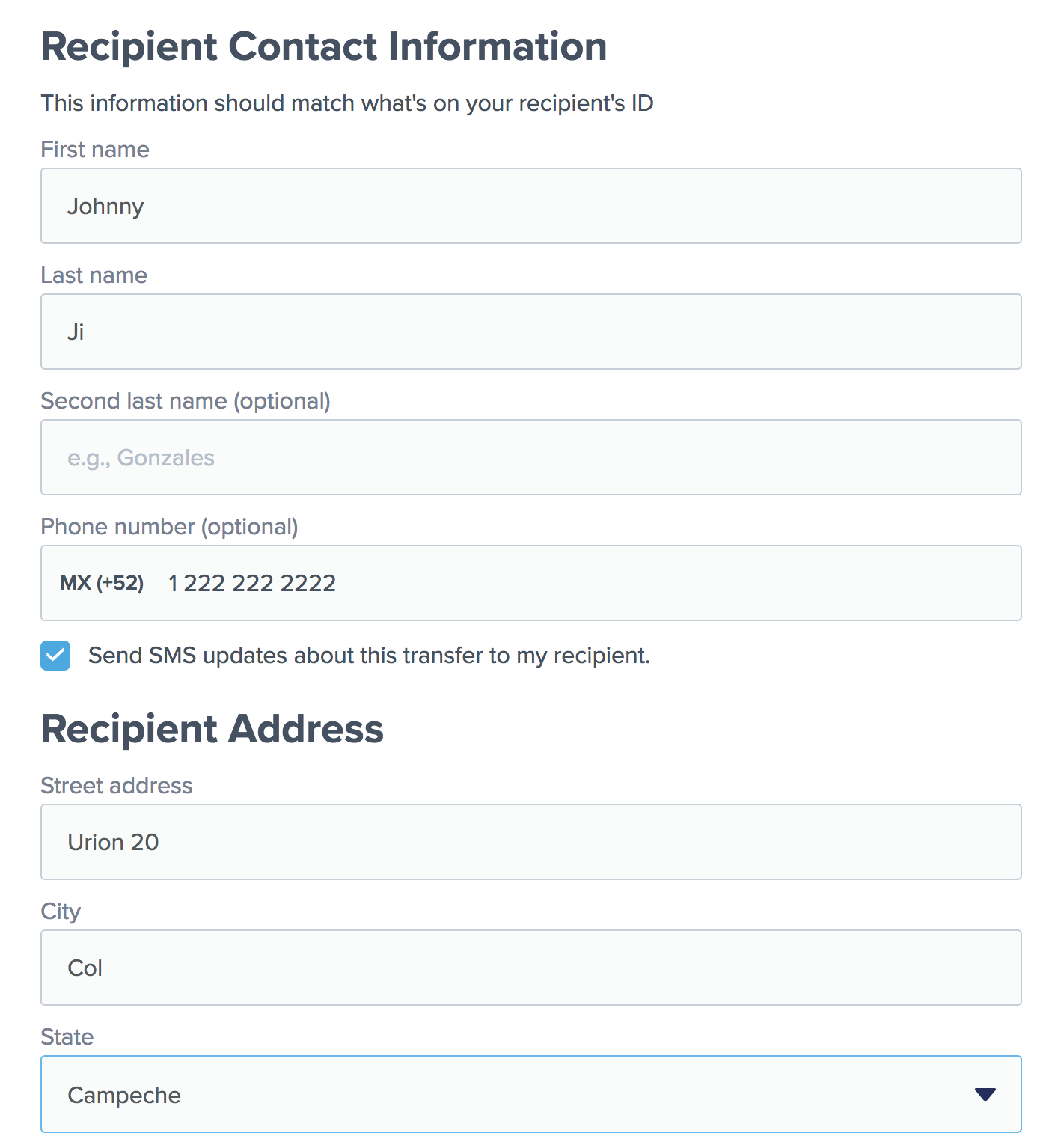
1. Register account with valid email address, password, and the recipient country.



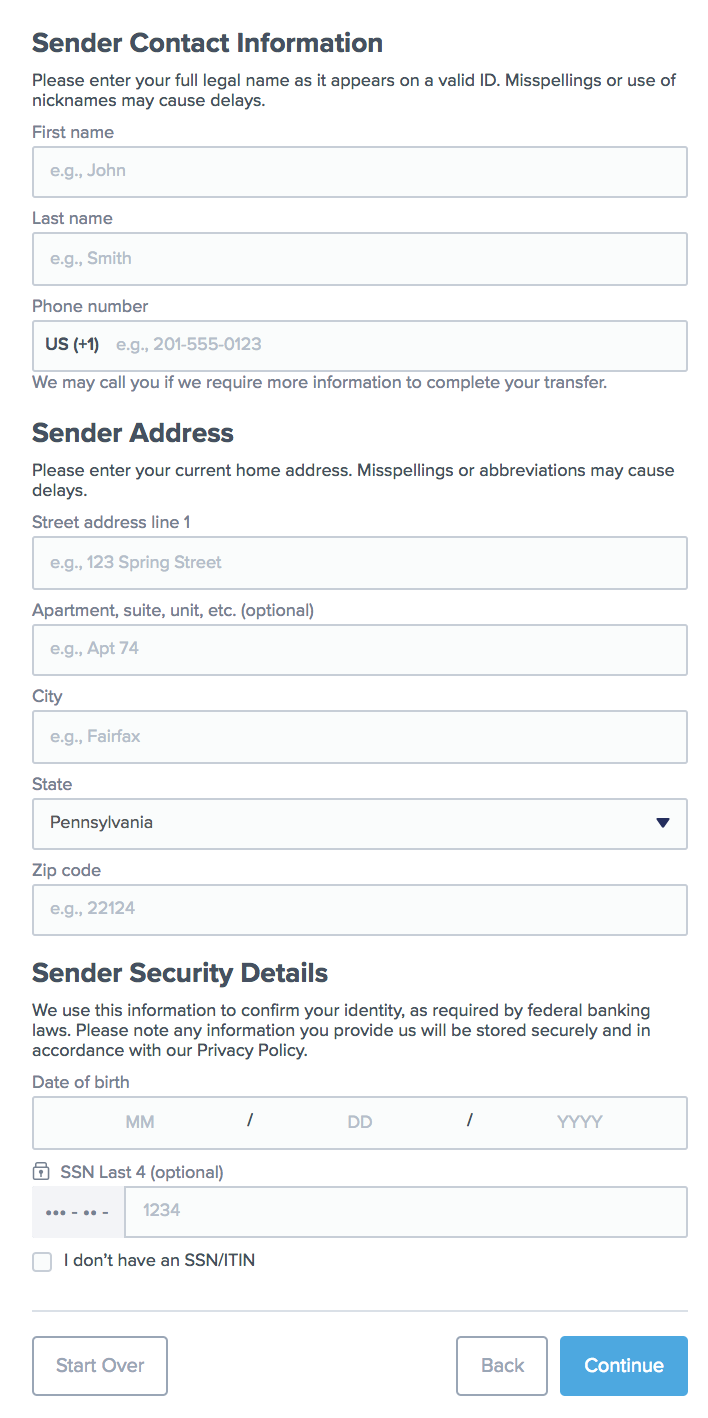
1. Input amount to transfer and currency. The exchange rate is then auto-populated.
2. Choose delivery speed and method. Here we chose Cash Pickup and Wal-Mart as the pickup location.



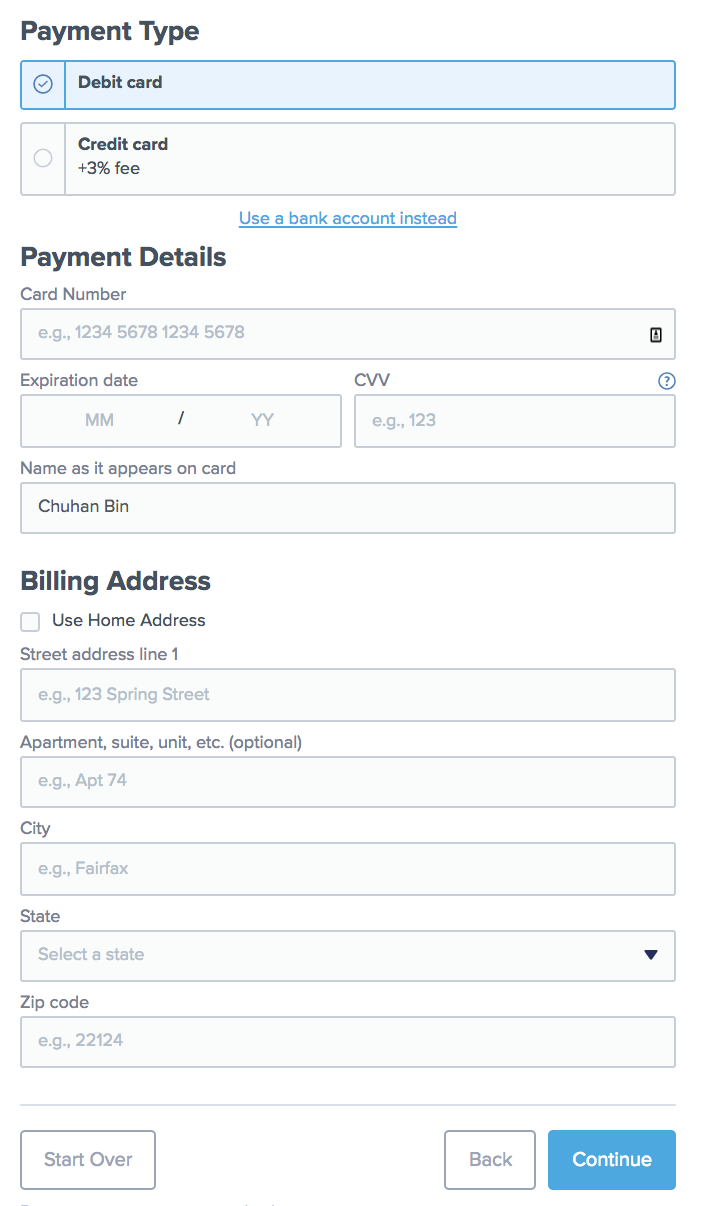
1. Input recipient information.



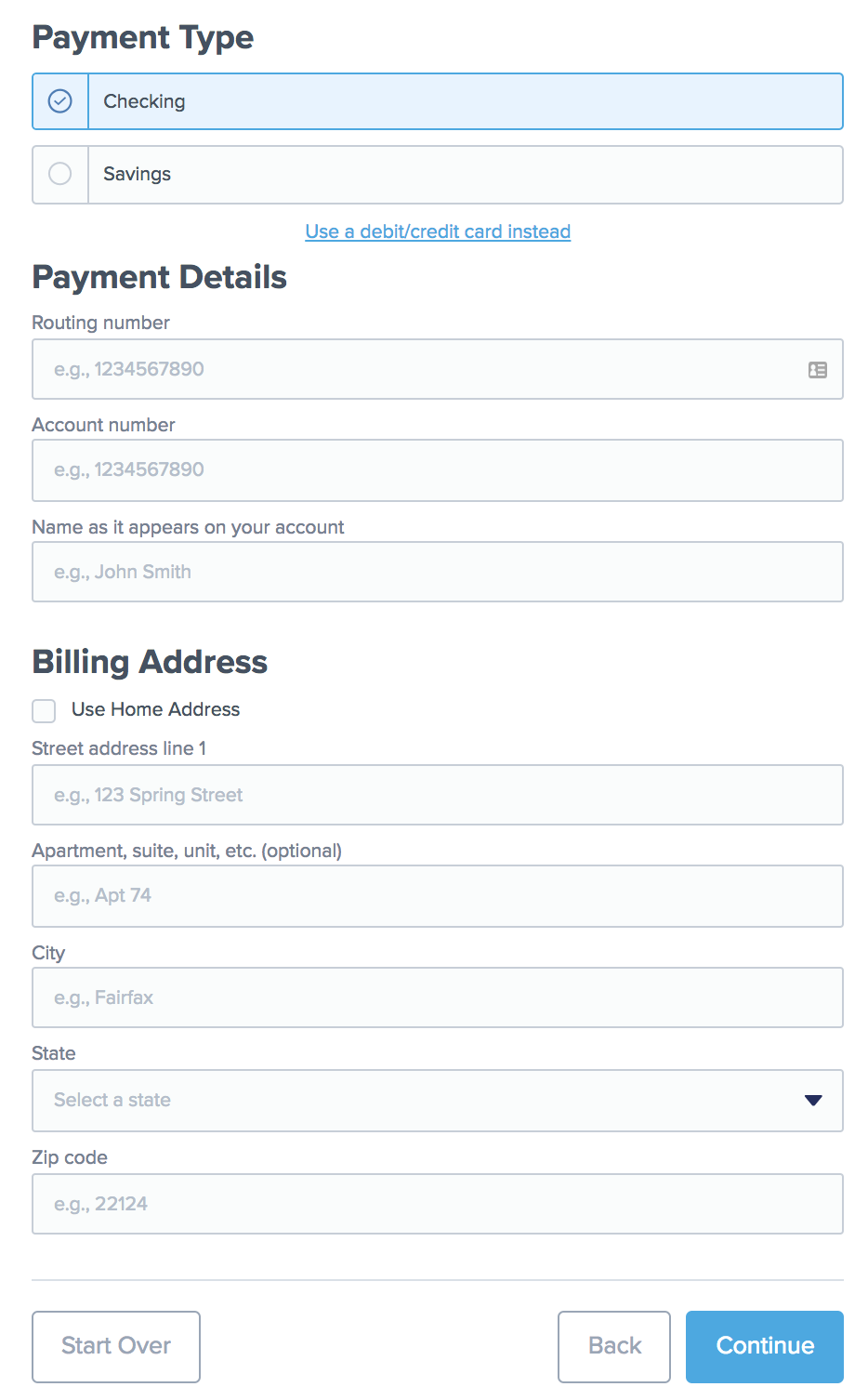
1. Input sender information. Remitly requires full legal name as it appears on a valid ID, date of birth, and optionally last 4 digits of SSN to confirm sender’s identity, as required by federal banking laws. Such information will be stored securely and in accordance with their Privacy Policy.



1. Input payment method.
   1. Debit/ credit card



* 1. Bank account



1. Confirm and send.

**Cash Pickup**

Remitly supports cash pick-up at hundreds of thousands of locations worldwide. In order to pick-up cash, recipients must:

1. Go to a cash pick-up location supported by the option the sender selected.
2. Bring a Photo ID, and the Remitly transaction reference number found on sender’s receipt (example: R1234567890).
3. Let the agent know they have a remittance through Remitly.

# Fraud types and algorithmic fraud detection logic

We identify the following major types of fraud that might occur. The four types are:

1. Money Laundering
2. Impersonation
3. Terror Financing
4. Elder Financial Abuse

Also, in this section we will introduce the first method in our fraud detection system using algorithmic fraud detection logic. This means we design rules in advance to classify different kinds of frauds. Our data inputs include data from both Wal-Mart and Remitly. From Wal-Mart side, our available data includes: the personal information and identify of the recipient, the photograph/CCTV cameras of the recipient, how much money is received by the recipient, when and where the money is received, and the information of the device the uses for verification. These information is available for all historical transactions. From Remitly side, our available data includes: location and personal credentials of the sender, the recipient’s personal information, the amount, currency and other details of the transaction. Also, these information is available for all historical transactions.

In our system, we make full use of these information to check if a transaction has the red flags identified. If a transaction meets the rules in our algorithmic fraud detection logic, we treat there are high possibility that such transaction is fraud. Our output for the algorithmic fraud detection logic is a vector of probabilities for each type of fraud. For instance, the vector of [100%, 30%, 0%, 0%] indicates this transaction has 100% probability on the fraud of money laundering, 30% of probability on impersonation and 0% of probability on other frauds. The detailed rules and algorithmic fraud detection logic is listed below for each type of frauds.

## Money Laundering

Remitly is potentially sensitive to money laundering (AML), which means moving large sums of money in many small amounts to avoid detection by financial authorities. The most common methodology AML compliance programs apply is the transaction monitoring framework aiming to detect suspicious activity. The scope of transactions is broad, covering deposits, withdrawals, fund transfers, purchases, merchant credits, payments, and others. Typically, monitoring starts with a rule-based system which scans customer transactions for red flags consistent with money laundering. When a matching pattern is observed, an alert is generated, and the case is referred to the bank’s internal investigation team for manual review.

Possible red flags and algorithmic fraud detection logic:

1. A series of transactions for large, round amounts being passed back and forth between same accounts
2. Multiple transactions made with different cards from different senders on a single IP address/on same device
3. A series of transactions for small amounts being sent from same sender in a short period of time, with the total amount to be quite large
4. A series of transactions for small amounts being received by same recipient within a short period of time, with the total amount to be quite large
5. High account turnover

For red flags 1-4, if such pattern is observed in a transaction, this indicates 80% of probability of having money laundering fraud. And if red flag 5 is observed, this indicates 50% of probability for money laundering fraud.

## Impersonation

An intruder might impersonate a legitimate sender (for example, by phishing the sender’s credentials, or the sender bank information leaked by untrustworthy parties) and send money to himself or a confederate. Although account holders are responsible for protecting their own private bank account information and card information, Remitly should also be able to use existing client data to detect anomaly.

Possible red flags and algorithmic fraud detection logic:

1. A series of transactions with unseen recipient based on sender’s transaction history with abnormally large amount
2. A recipient picks up money at a Wal-Mart at the location which is not the residential country/state for the recipient
3. Sender chose a recipient country that is unseen in the past transaction history.
4. The name of the sender is not the same as the name of the internet-account owner sending the payment
5. The country of the sender does not coincide with the country of the Internet account owner sending the payment
6. Payment made late at night, according to the local time of the sender or the recipient.
7. Sender/recipient credentials used almost at the same time at different locations

For red flags 4 and 7, if such pattern is observed in a transaction, this indicates 80% of probability of having impersonation fraud. And if other red flags are observed, this indicates 50% of probability for impersonation fraud.

## Terror Financing

Potentially sensitive to terror financing (CTF) concerns. Banks have long used anti-money laundering systems to flag suspicious activity, and in the aftermath of September 11st, they have turned to those same legacy tools to catch terror-related transactions. However, such legacy tools are not effective. One reason that ISIS so hard to fight is that the terrorist network is diffuse and scattered, with small cells of operatives all over the world, hence the pattern is much harder to be detected.

Possible red flags and algorithmic fraud detection logic:

1. Sender or recipient country is located in ISIS identified area and unseen in the past transaction history
2. Sender or recipient is associated with person included in the consolidated list of persons and/or entities issued by organization such as the United Nations Security Council Committee on Afghanistan.
3. Transaction with large amount soon after new account opening

For red flags 1 and 2, if such pattern is observed in a transaction, this indicates 80% of probability of having terror financing fraud. And if red flag 3 is observed, this indicates 30% of probability for terror financing fraud.

## Elder Financial Abuse

Elder Financial Abuse is also a big issue in fund remittance. Seniors are often deliberately targeted by scams and fraud. Persons over age 50 in the US control over 70% of the nation's wealth. One big source of international transfers is immigrants/ foreign workers sending money back to their home country to support their aged parents. Elderly are more likely to have disabilities that make them dependent on others for help. These "helpers" may have access to homes and assets and may exercise significant influence over them. In addition, advances in technology have made managing finances too complicated for seniors to understand. Such cases also apply to females who received limited education.

Although elder financial abuse is common in financial crimes, it is also easier to detect comparing to AML and terror financing.

Possible red flags and algorithmic fraud detection logic:

1. Large, unexplained transfer-outs in the older person's Remitly account
2. CCTV cameras in Wal-Mart stores capture suspicious persons accompany/ following elderly recipient
3. If the recipient information is on the blacklist of common bank accounts committing elder financial abuse

For red flags 2 and 3, if such pattern is observed in a transaction, this indicates 90% of probability of having elder financing abuse fraud. And if red flag 1 is observed, this indicates 50% of probability for elder financing abuse fraud.

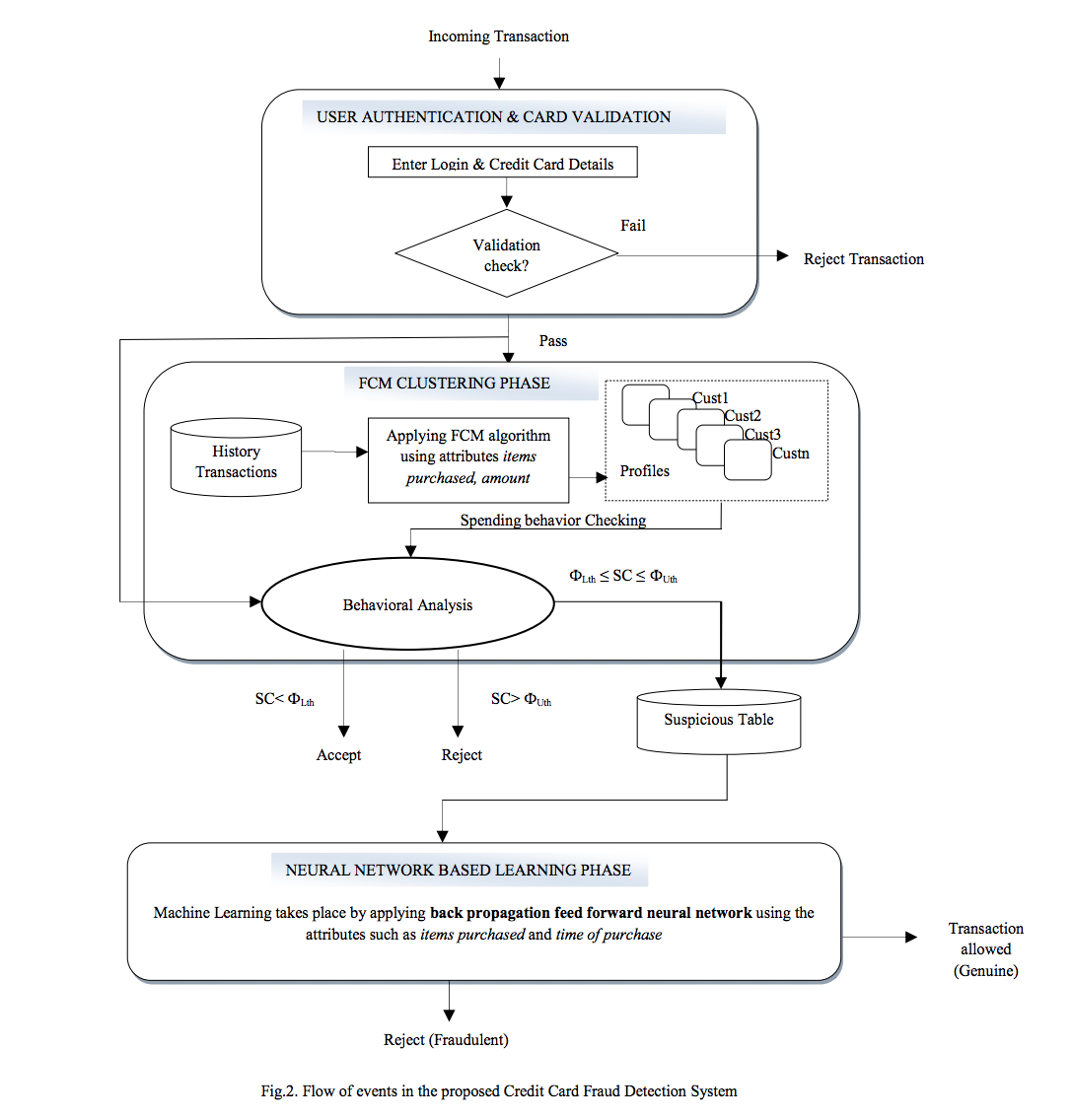
However, the algorithmic fraud detection logic also has some drawbacks. The false positive rates generated by this rule-based system can reach very high levels, sometimes 90% or more. And since the system is rule-based, it is rigid and likely fails to consider complex interactions between the various behaviors used for financial frauds. Also, the rules are pre-identified, this means the rule-based system will fail to detect new patterns in frauds. Furthermore, maintaining this program is expensive, partially due to the cost of assigning a large team of investigators to the task. This is the reason that we will introduce another machine learning approach in the next section.

# Machine Learning Algorithms

In addition to the traditional rule-based system, automated machine learning is a modernized tool by which Artificial Intelligence (AI) is used to select the best machine learning algorithms for making predictions from a particular dataset. All of the information being passed to the reviewer can instead first be assembled into a modeling dataset and passed into a set of machine learning models. An AI algorithm can then guide the process of the model selection process which most accurately capture fraud. Once the best model is identified, it can be applied to the validation cases–a random sample from the modeling dataset that was not used to train the model. The cases can be scored, and the cut-off score that captures all of the suspicious activities in this historical dataset is noted for use in deployment.

Machine learning algorithms are increasingly used to mine vast quantities of bank data and find anomalies in accounts and transactions that might otherwise have gone unnoticed. Such ML models are based on a combination of factors, including how quickly money moves around, where it's moving, and how much is being transferred. They also look for clues like anomalies in invoicing number sequences. If a criminal group is looking to launder money, it might falsify invoices to make it appear a legitimate transaction occurred, when, in fact, the money came from a drug deal or the sale of counterfeit goods. ML models can spot duplications and mistakes in the system and reveal the underneath pattern.

Machine learning algorithm has long histories to work along with financial system back to 20 years ago for fraud detection. Currently, more and more neural network model implemented in fraud detection system. As the figure below shows, in logistics, this algorithm is mainly the bottom block to save huge human cost and reduce human mistake.



In this report, multiple neural networks are under consideration, not limited to simple back-propagation. There are 3 different topologies to be evaluated according to speed, memory, and accuracy: Artificial Neural Network (ANN), Adaptive Neuro-Fuzzy Approach (ANFIS), and Self-Organizing Mapping Neural Network (SOMNN). We will give a brief introduction to these algorithms and analyze the pros and cons for each algorithm.

As a basic setting for our machine learning algorithms, the data input is from two sources (Wal-Mart and Remitly) similarly as mentioned in the previous section. In our models, we assume that we don’t have accurate and complete label of the probability vector in the train data. Thus, our model supports unsupervised learning. The output of our machine learning algorithm is also the probability vector as mentioned in the previous section.

## Artificial Neural Network Back-Propagation Algorithm

Artificial Neural Network is one of the most basic neural network learning methods. It could process complex and huge input data into a classification outcome efficiently and easily to implement. It has only three components: output layer, the hidden layer, and input layer, with weighting factors, step function, summation functions, active function, and learning function.

The classical back-propagation (BP) algorithm is a common learning algorithm in supervised training situation in classification. The following figure (Figure 1.1) shows the detailed flow through the algorithm.

Also, back-propagation can be used in unsupervised learning with the auto encoder algorithm. In the unsupervised auto encoder algorithm, no label is required (which might be the case for Remitly and Wal-Mart), this algorithm use the original input as the output and has an encoding process which compress input data into dense data and then, has a decoding process which decode the compressed data and return the original data as output. The auto encoder algorithm can be used for unsupervised classification and anomaly detection, which perfectly fits our problem here. We can first train the auto encoder neural network on certain data and fit the model on new data, if the new data outputs a result similar to itself, this means the trained model fits the new data and the new data falls into the same classes as the trained data. Otherwise, it means the new data belong to a new different class.

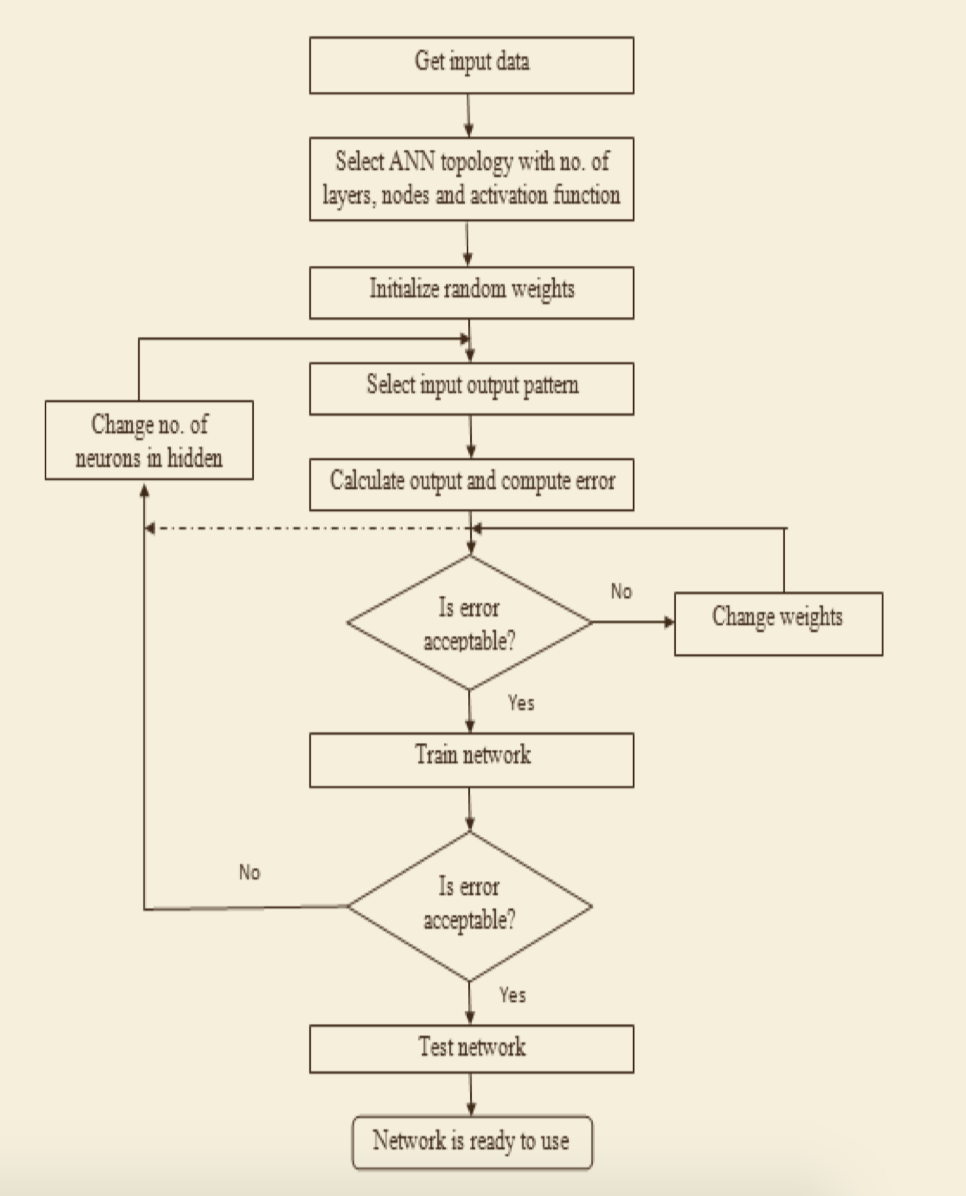


Figure 1.1. flow of back-propagation

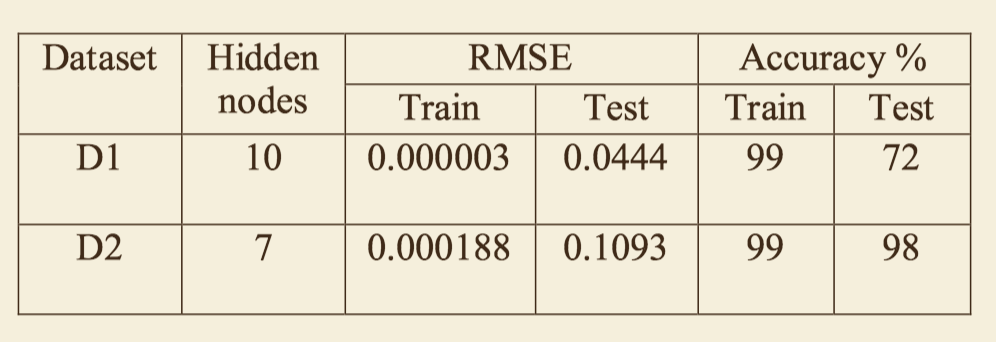
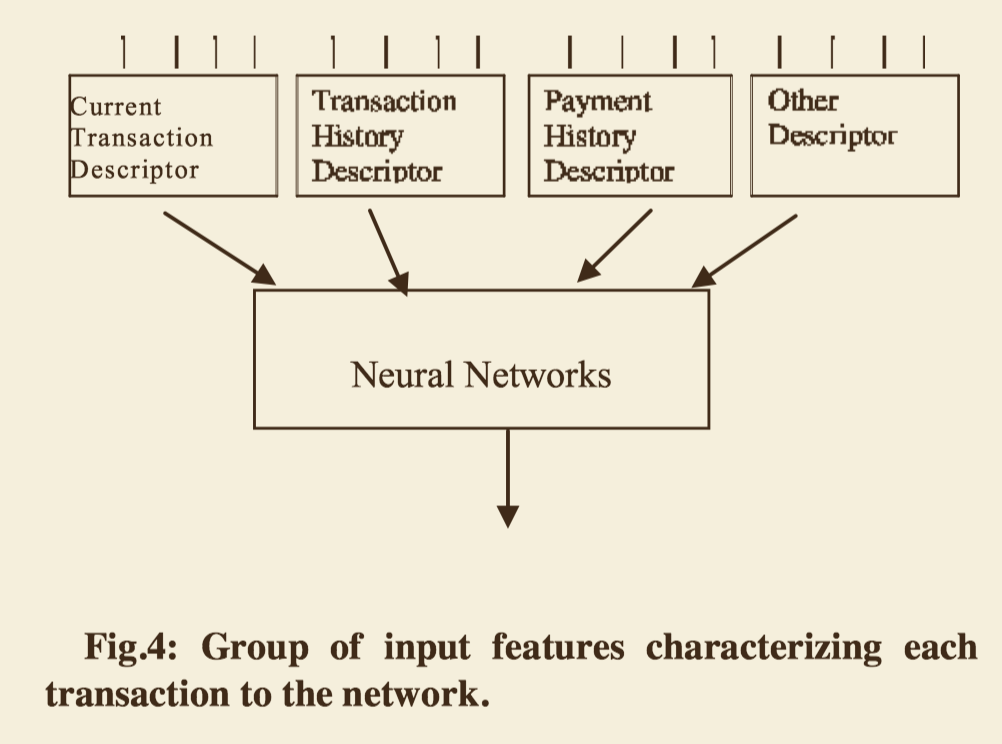


Figure 1.2. left: input data right: dataset characteristics

When sending the data request, for example, in figure 2, the network need information about current transaction description, transaction history description, payment history description and other description. All the input data is very common and easy to fund in the banking system.

From the experiment by Prianka Patil couple months ago (Patil, 2018), the highest accuracy rises up to 98% and 72% with 300 user records with different dataset.

**Pros and Cons**

Compared to traditional machine learning, ANN is good to handle large dataset with efficient cross-validation, when data is GB and TB large the network still works very stable and accurate, particularly for non-linear input data.

It can achieve high computing speed with the implemented in GPU and TPU where the service is very easy to use in AWS, Azure or Google Cloud. Usually, the data training with CPU in 3 days will reduce to 1 days with GPU support, and the accuracy is much higher than common machine algorithm, such as KNN, DT, even random forest.

Traditional machine learning is feature learning, however, ANN is representing learning widely used in classification and forecasting problem. Simply, this means engineer does no longer need to select related feature in feature engineering. The network will progress itself to compute the data in order to finish the classification.

However, ANN algorithm may still be a black box to most engineers who implement it into the system. So, the banking system may require high-skill professional to set up the system and the labor cost might be high and hard to debug.

Another thing to pay attention is that ANN need to retrain every time, and the weights factors need to change depends on the error from the back-propagation algorithm.

## Self-organizing Mapping Neural Network

Compared to the standard ANN algorithm, SOMNN is a modified optimization algorithm to improve higher accurate of fraud detection. It is a learning method to divide a set of input pattern into a cluster which is inherent to the input data.

In the self-organized mapping neural network (SOMNN), instead of just computation architecture, dataset description and number of the data point are all included in SOMNN engine. This system can determine when a transaction is to be processed, blocked, unblocked or alert set off. The biggest difference compared to ANN, SOMNN can construct the fraud detection and determine whether the transaction is legitimate or fraudulent and give the reason of any alert right after the transaction. All fraudulent record will be immediately terminated.

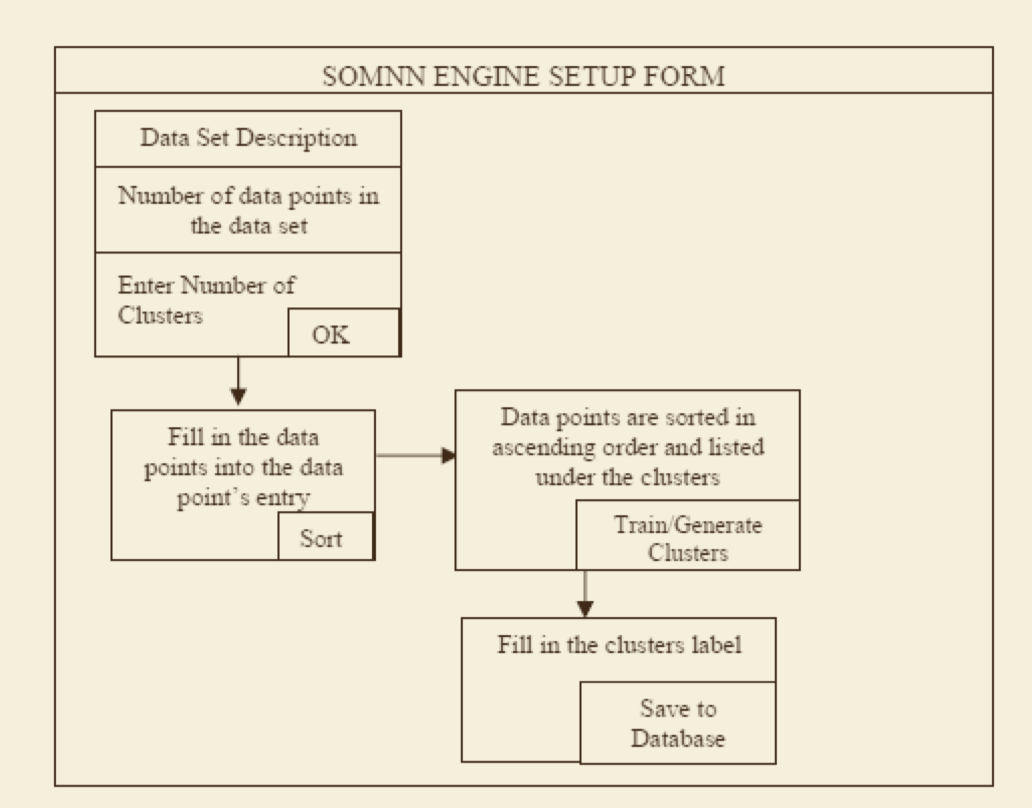


Figure 2.1 SOMNN setup form

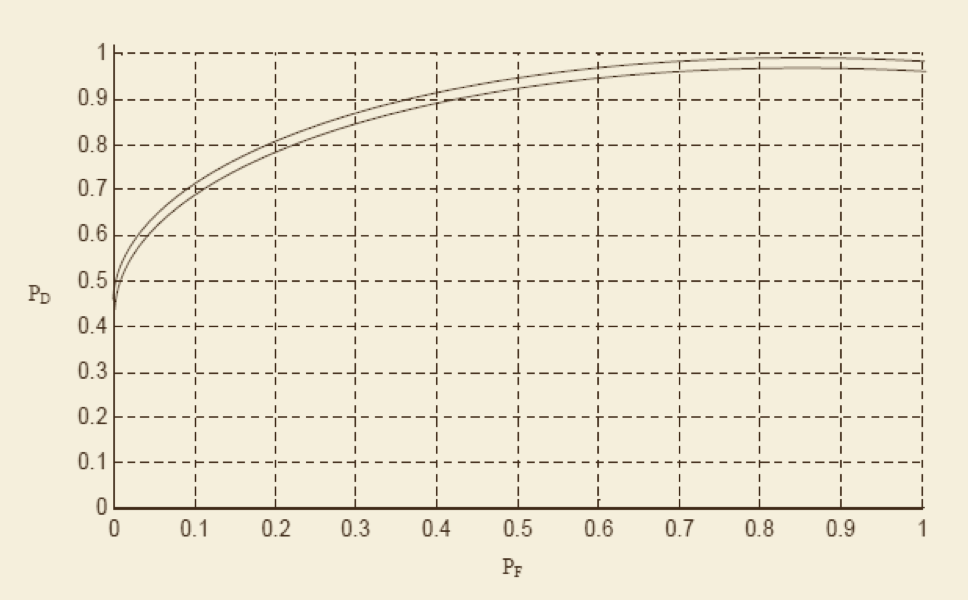
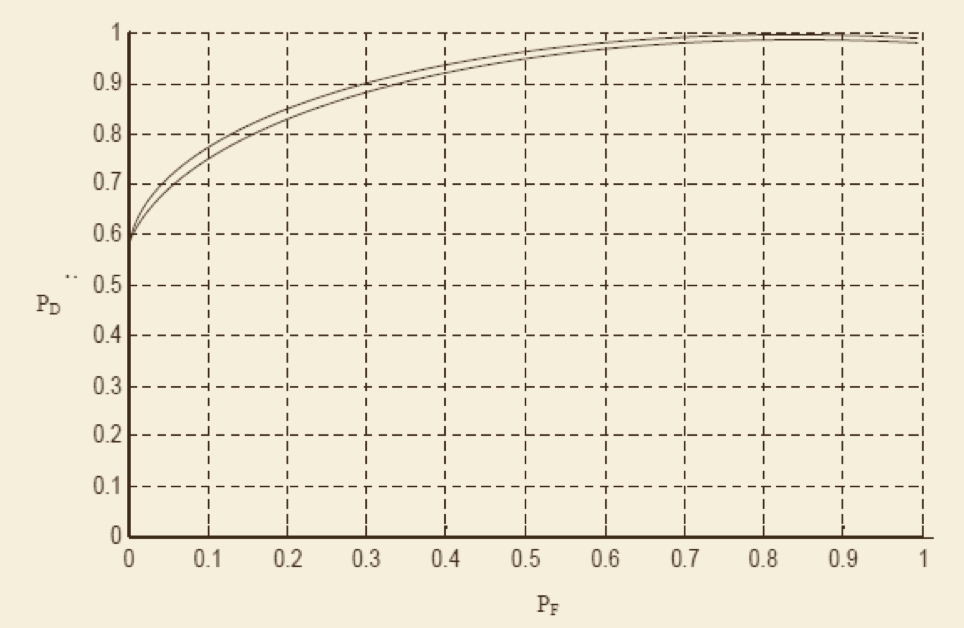


Figure 2.2 left: Receiver-operating curve real fraud probability vs detected fraud probability in withdraw

right: Receiver-operating curve real fraud probability vs detected fraud probability in deposit

**Pros and cons**

SOMNN constructs multiple clusters rather than many ordinary inputs. This will improve the model performance in term of accuracy and stability. As the above outcome summary shown figure 2.2, in both of withdrawal and deposit operation, the Receiver-operating curve (ROC) curve gives conservative detection which protects Remity and its clients.

On the other hand, this reliable detection leed 95% accuracy to correctly point out the fraud without causing a false alert. Remity may still need to deal with the rest 5% which as a type II error. Even it is unharmed, time and cost still may happen.

## Adaptive Neuro Fuzzy Approach

Adaptive neuro fuzzy approach or adaptive neuro fuzzy inference system (ANFIS) is widely used as the demand of self-learning predictive system rise. ANFIS is a hybrid neural network system combined with fuzzy inference ---- this leads the system can automatically learn from the newer instance of fraud.

In ANFIS, both numeric and linguistic rules can be easily combined together. Any previously given rule or prior knowledge formatted as linguistic rules will help the system training faster. The prior rules can be feed into the system as part of the input to the inference engine; furthermore, ANFIS may automatically generate new rules based on the permutation and combination of all the inputs. Due to this reinforcement system, the base rule will be updated as every single successful transaction and fraudulent transactions happened.

The following figure explores a detailed architecture of the ANFIS system to implement.

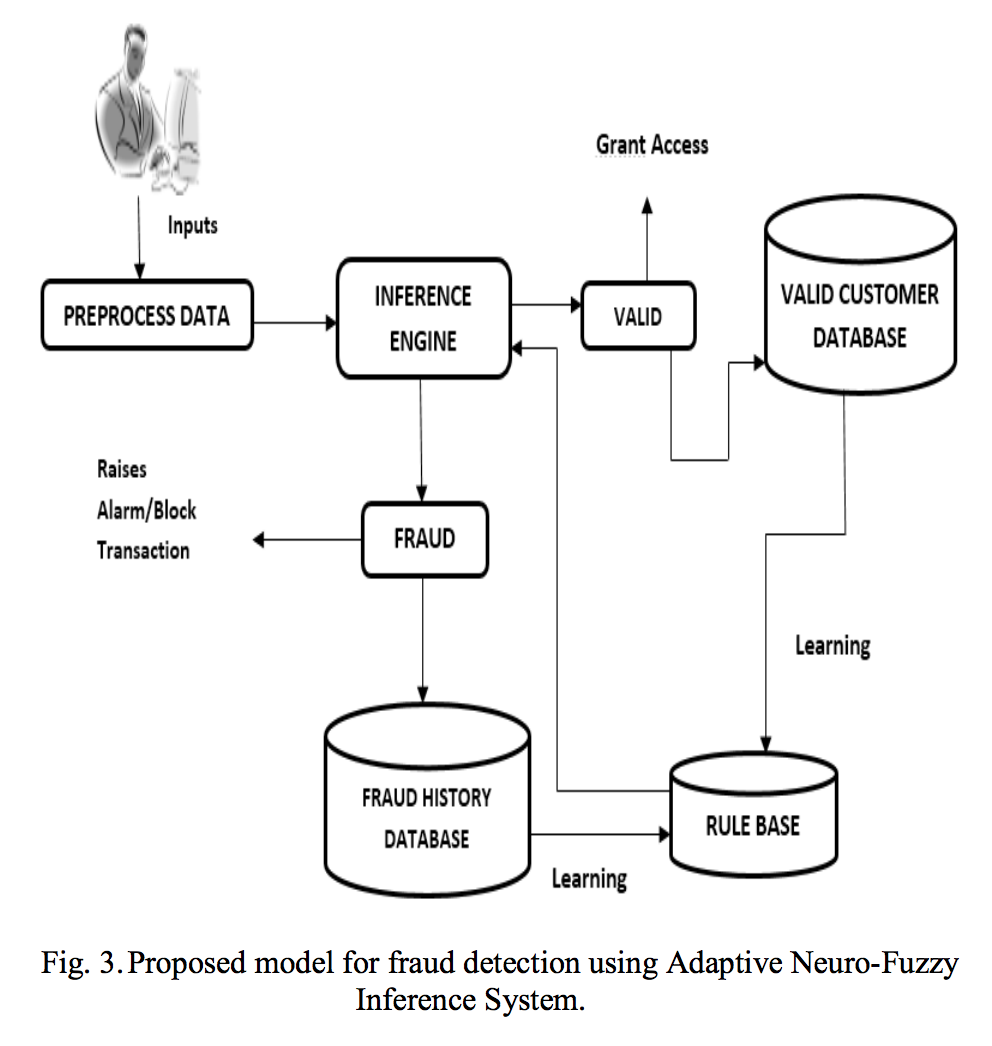


Figure 3.1 model of ANFIS

The experiment is done by Shaji last year (Shaji, 2017), the below table summarizes the accuracy of ANFIS, ANN and Bayesian Network under the same data and training environments. ANFIS achieves the lowest error rate in RMSE, MSE and MAE also takes less time to train.

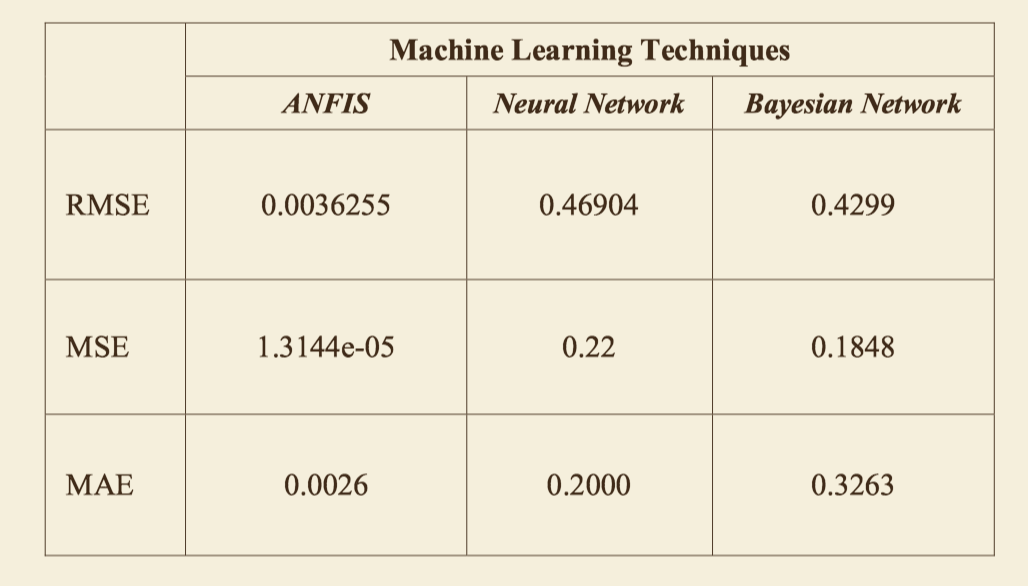


Figure 3.2 result of different machine learning algorithm

**Pros and cons**

This system allows Remity to combine both of the benefits from generating fuzzy rules and inference, as well as benefits from self-learning neural network. The high accuracy may be caused by the combination of input features and re-reinforcement learning the new data over and over again. Besides that, all the weights or bias factors will be automatically updated in this self-learning system.

The initial training may take some time and effort, similar to ANN; however, once the ANFIS system has been well trained, not need to retrain or change the weights factors. It can update and adjust itself.

ANFIS may need input data formatted in matrix form as this network required.

# System Design and Conclusion

To conclude, in our fraud detection system, we have identified four different kinds of fraud: money laundering, impersonation, terror financing and elder financial abuse. In our system, we apply two methods for fraud detection: the first is the rule-based algorithmic fraud detection logic, and the second is the machine learning algorithms. Especially, the second machine learning algorithms can compensate the rule-based system on its disadvantages of relying on pre-identified logics and failure to analyze new patterns, having lower accuracy due to the high false positive rates and higher labor cost for maintaining and updating the rules.

For both of the two methods, our system will generate an output of probability vectors, indicating the probability of the transaction falling into the identified fraud. Our final result of the system will combine both the result from two methods by taking average. For instance, for a specific transaction, the first method outputs a vector of [80%, 50%, 0%, 0%] and the second method outputs a vector of [60%, 50%, 40%, 0%]. The final output of our system will be [70%, 50%, 20%, 0%], meaning that this transaction has 70% likelihood of money laundering fraud, 50% likelihood of impersonation fraud, 20% likelihood of terror financing fraud and 0% likelihood of elder financial abuse fraud. Our system will report this fraud detection result to Wal-Mart and Remitly. Then, Wal-Mart and Remitly will decide how to act on these results to minimize fraud in the financial process.

# References

Faggella, Daniel. “Machine Learning in Finance - Present and Future Applications.” TechEmergence, 1 June 2018, [www.techemergence.com/machine-learning-in-finance/](http://www.techemergence.com/machine-learning-in-finance/).

Credit Card Fraud Detection with a Neural-Network. www.bing.com/cr?IG=CDA9AC6DBEC644D4A59E0185AED56955&CID=3634FDEE71F86C0F09D1F1FA70056D73&rd=1&h=SA0FSVUKPaTIikUjcd3P3t9UKUnI73nl66gqGkADdHY&v=1&r=http://bit.csc.lsu.edu/~jianhua/quang.pdf&p=DevEx.LB.1,5498.1.

“Credit Card Fraud Detection: A Hybrid Approach Using Fuzzy Clustering & Neural Network - IEEE Conference Publication.” Design and Implementation of Autonomous Vehicle Valet Parking System - IEEE Conference Publication, Wiley-IEEE Press, ieeexplore.ieee.org/document/7306735/.

Akhilomen, John. “Data Mining Application for Cyber Credit-Card Fraud Detection System.” SpringerLink, Springer, Dordrecht, 16 July 2013, link.springer.com/chapter/10.1007/978-3-642-39736-3\_17.

Akhilomen, John. “Data Mining Application for Cyber Credit-Card Fraud Detection System.” SpringerLink, Springer, Dordrecht, 16 July 2013, link.springer.com/chapter/10.1007/978-3-642-39736-3\_17.

Five Steps Towards Effective Fraud Management. [www.bing.com/cr?IG=A50287DB842C43BA8D80A9C5D06881B7&CID=07CEF3C558766C782FA2FFD059C16DD5&rd=1&h=mYS8rcUWRXOLtSQdz08tHQ9siXiiJUGVlc2Jkma\_TEg&v=1&r=https://www.firstdata.com/downloads/thought-leadership/ecommerce-fraud-best-prac.pdf&p=DevEx.LB.1,5508.1](http://www.bing.com/cr?IG=A50287DB842C43BA8D80A9C5D06881B7&CID=07CEF3C558766C782FA2FFD059C16DD5&rd=1&h=mYS8rcUWRXOLtSQdz08tHQ9siXiiJUGVlc2Jkma_TEg&v=1&r=https://www.firstdata.com/downloads/thought-leadership/ecommerce-fraud-best-prac.pdf&p=DevEx.LB.1,5508.1).

<https://blog.datarobot.com/anti-money-laundering-using-automated-machine-learning>

<https://www.wired.com/story/quantaverse-ai-terrorist-funding/>