Smarter STIP

In Oct 2020, Visa successfully launched Smarter STIP, the first real time decisioning deep learning model in Visa. It determines authorization decisions when issuing banks are not available to do so, and the decisions are learned from the banks' own behavior. Since its global launch in October 2020, there have been hundreds of active clients signed up for the service and ongoing active client engagement involves clients from every Visa region. Smarter STIP is the first product from the Visanet +AI initiative, which aims at solving pain points for financial institutions, processors, merchants and FinTech companies through Artificial Intelligence powered solutions and Visa's existing network connections. Besides Smarter STIP, other products such as Smarter Posting, Smarter Settlement Forecast and Smarter Account Verification are either in production or under development. The successful productionize of the Smarter STIP model demonstrates the value Artificial intelligence could bring to Visa and the payment industry. This article captures the details of the analytical innovation and learnings.

Figure 1: The Wall Street Journal article featuring Smarter STIP

Background

Stand-in-Processing (STIP) is initiated when an issuing bank is unable to provide a real-time response to Visa, due to planned or unplanned outages or network problems. Traditionally, when these events occur, Visa "stands in" on behalf of the issuer to approve or decline a transaction through our hosted STIP process using authorization parameter (pre-defined rules) set by the issuer. The following diagram shows the flow of the traditional STIP process.
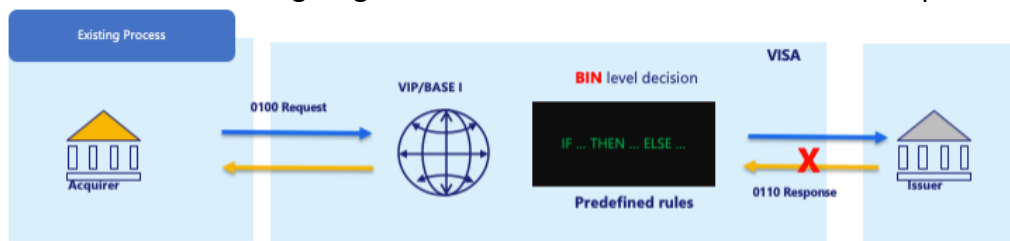

Figure 2: Rule based STIP process for transactions decisioning

Limitations of rule based STIP:
- BIN level based: all accounts with the same issuer BIN are treated the same regardless of history
- Static predefined rules: not able to dynamically adjust strategy

- Since issuing banks are not available, they tend to be overly conservative when defining the static rules, which results in more than 25% lower approval rate during STIP incidents
    - Influential STIP approval rate is only **67%**
    - Issuer approval rate on average when they are available is **93%**

Opportunities and Motivations:
The opportunity to enhance the logic for decisioning with an AI-driven solution was evident from observed historical incidents. The following were two incidents with large final impact:
- On 02/15/2019, a large Canada-based processor experienced 8 hours of STIP outage. All 600 k transactions were declined with 0% approval rate.
- On 01/17/2019, one of the largest banks in Latin America experienced 3 hours of STIP outage. 200 k transactions were declined. The approval rate is only 49%.

The total addressable market is 101M outage transactions per year.

Building the Smarter STIP model
In order to make the STIP process "smarter", we developed a machine learning model that could "learn" the issuer's strategies and logics for transaction decisioning, and then "mimic" these strategies during outages - making approval/decline decisions as close as possible to what the issuer would have done. It is worth noting that Smarter STIP is not a fraud model, since issuers take into consideration elements other than fraud (e.g. credit risk, account balance) when making an authorization decision.   The table below highlights the differences.

	Fraud Model	Smarter STIP Model
Target	Prevent Fraud	Predict Approval or Decline when issuers are offline - Only a small portion of declines are due to suspicious fraud - Majority of the declined transactions are due to factors like activity limit or not enough balance
Criteria	Identify fraudulent transactions	Mimic issuers' strategy

Table 1: Differences between a model targeting fraud and Smarter STIP

The team experimented with different algorithms, including traditional machine learning models like gradient boosting trees and random forest. Using deep learning techniques resulted in a 10% improvement over traditional machine learning algorithms; and using a multi-layer RNN model with LSTM cells had the best performance. The following graph shows the overall architecture of Smarter STIP.
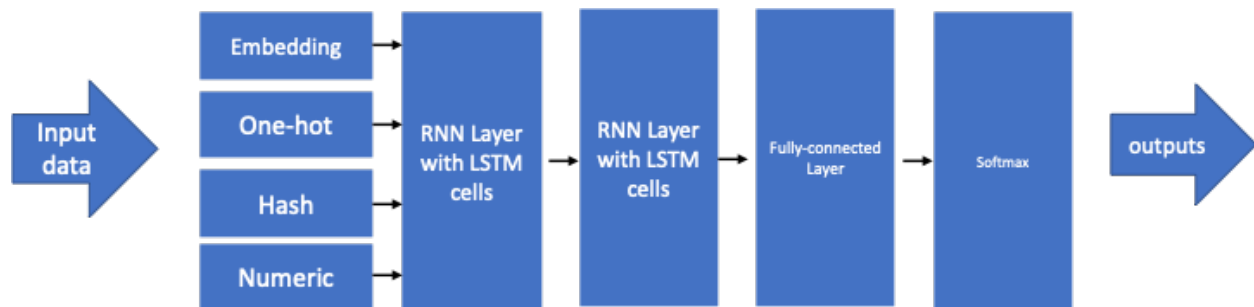
Figure 3: Smarter STIP deep learning model architecture

The Smarter STIP model was trained on six months of transactions with billions of transactions and 2.9 million parameters. The STIP model has 6 different geographical segments: US, Canada, European Union, Asia Pacific, CEMEA (Central Europe, Middle East, and Africa) and Latin America. The models are able to achieve over 95% match rate when compared with issuers' decisioning.

Technical Innovations

In order to achieve high accuracy and low latency results, the team had to innovate, below are some of the highlights.

- Leverage **NLP (Natural Language Processing) techniques** to model Visa transactions. Traditionally transactions are treated separately as individual records. We treat all the historical transactions with the same account as one training example. This approach enables the model to learn the long history of each account as a group, which provides great insights on how the issuers make their own approval/decline decisions. This innovation significantly improves the model accuracy.
- Utilize **Transfer learning** to take advantage of large amount of Visa transaction history. If we focus only on the STIP transactions, there is not enough data and no reliable labels to enable a successful outcome. Instead, we take advantage of billions of non-STIP transactions to enhance the model performance.
- The Smarter STIP model takes more than one week to train, and thus traditional feature importance algorithm is not feasible for this model. To address this issue, we developed an efficient algorithm to calculate **feature importance for Deep Learning models**. This algorithm not only significantly reduces the time of the feature selection process, but also reduced the deployment effort and inference latency.
- Since there is no universal optimal solution like fraud prevention, we need to make sure the model can mimic issuers' strategies for more than 17K BINs across all the regions. Regular transaction-level sampling strategy does not work since we need to consider all

historical transactions from one PAN. Instead, we developed an efficient sampling algorithm at the PAN level to balance the distribution for all the individual BINs.

In order to deploy the model, the team leveraged Visa's AI Platform (AIP).   As part of this effort, a call-out had to be developed between Visa's transaction system and AIP to retrieve the score. This in itself required significant technology innovations in order to enable high throughput and low latency. Lower latency was achieved by tweaking the design of the model.

Model Stability

As with any model development, it is important that a model is robust. A robust model is one that can withstand the test of time. Some models might appear very accurate during development, but as time goes by, they quickly deteriorate. In order to measure for robustness, one has to perform a simulation over time to assess the potential deterioration.

For Smarter STIP, a window of six consecutive months was used to test for robustness. Figure 4 depicts how the F1 score[2] (left y-axis) remains stable throughout a period of six months. The consistent pattern indicates score stability and model robustness.   In other words, this means that the score is able to closely mimic issuers' strategies for an extended period of time. We achieve similar results for accuracy, precision, and recall[2]. Those additional metrics are not included in the graph for simplicity.

The orange bar (right y-axis) shows the difference between the model approval rate and the issuer approval rate. The differences are within 1% on average. As the model runs for longer time, the difference is even smaller, indicating the model **"learns"** more of the issuers' strategies. This is a characteristic of these type of models, they continue to learn as time goes on.
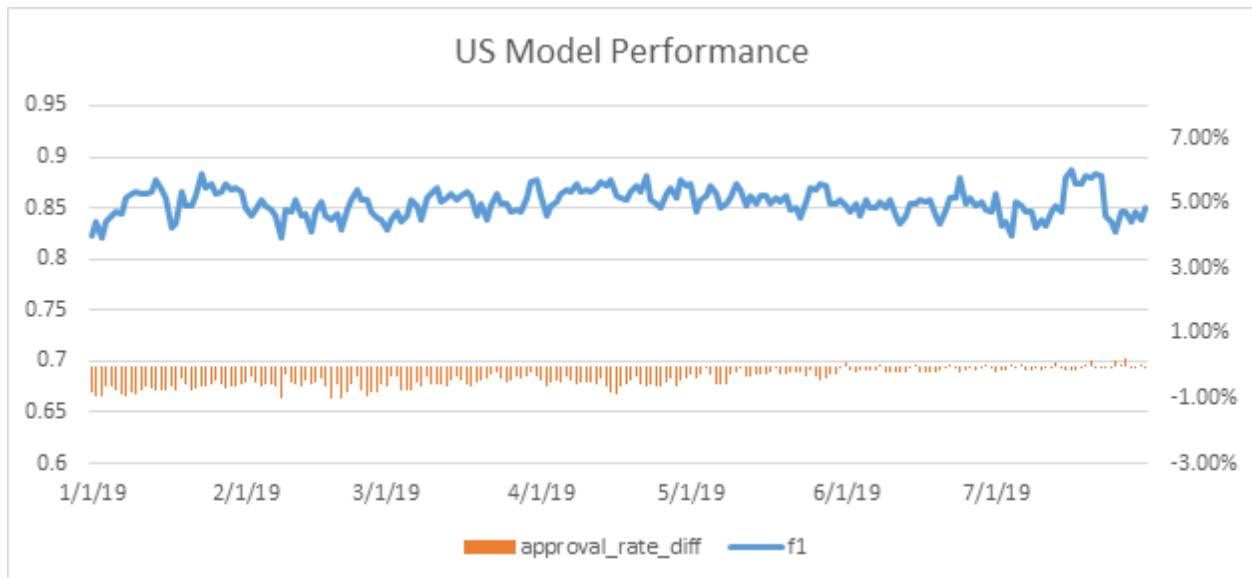

Figure 4 Smarter STIP US model robustness assessment

Business Impact

We have started to see the benefits of this model in production such as an increase in the approval rates of transactions that are legitimate and low risk. This was experienced with a recent outage from a European Bank, where we saw an increase in approval rate by more than 30% for card-not-present and 20% for F2F transactions due to Smarter STIP decisioning.

UBS	Issuer (Typical)	Smarter STIP (Oct'20 – Nov'20)	STIP before Oct'20 (Oct'19 – Sep'20)
F2F	98.5%	98.1%	77.9%
CNP	89.8%	93.1%	59.4%

Table 2 Statistics on real outage. Comparing with the STIP approval rate before Oct 2020, Smarter STIP significantly improved the approval rate for UBS, and it's much closer to the approval rate when the bank is available.

Conclusion

This effort is the result of a collaboration of several teams in both Technology and Product bringing the following benefits:

- Reduces cardholder friction during issuer outages
- Serves as the playbook for other real-time model deployments in VIP. Teams had to solve issues with scalability, transaction burst-out, latency, etc.
- Multiple technical innovations are developed in this journey, and all future AI models will benefit from these innovations
- Capabilities in place to deploy self-learning models that leverage real-time streaming data and historical data
- First model to test and validate the call-out between AIP and VIP which allows machine learning scientists to leverage more sophisticated algorithms, therefore achieving higher accuracy and coverage.
- Semi-automated model onboarding process to reduce manual work and accelerate time to market.

Reference:
1. WSJ Report: https://www.wsj.com/articles/visa-unveils-more-powerful-ai-tool-that-approves-or-denies-card-transactions-11598436001
2. Definition for F1 score, precision, recall, accuracy: https://en.wikipedia.org/wiki/F-score