

Product Whitepaper

SentiLink Synthetic Score: Pinpointing Synthetic Fraud

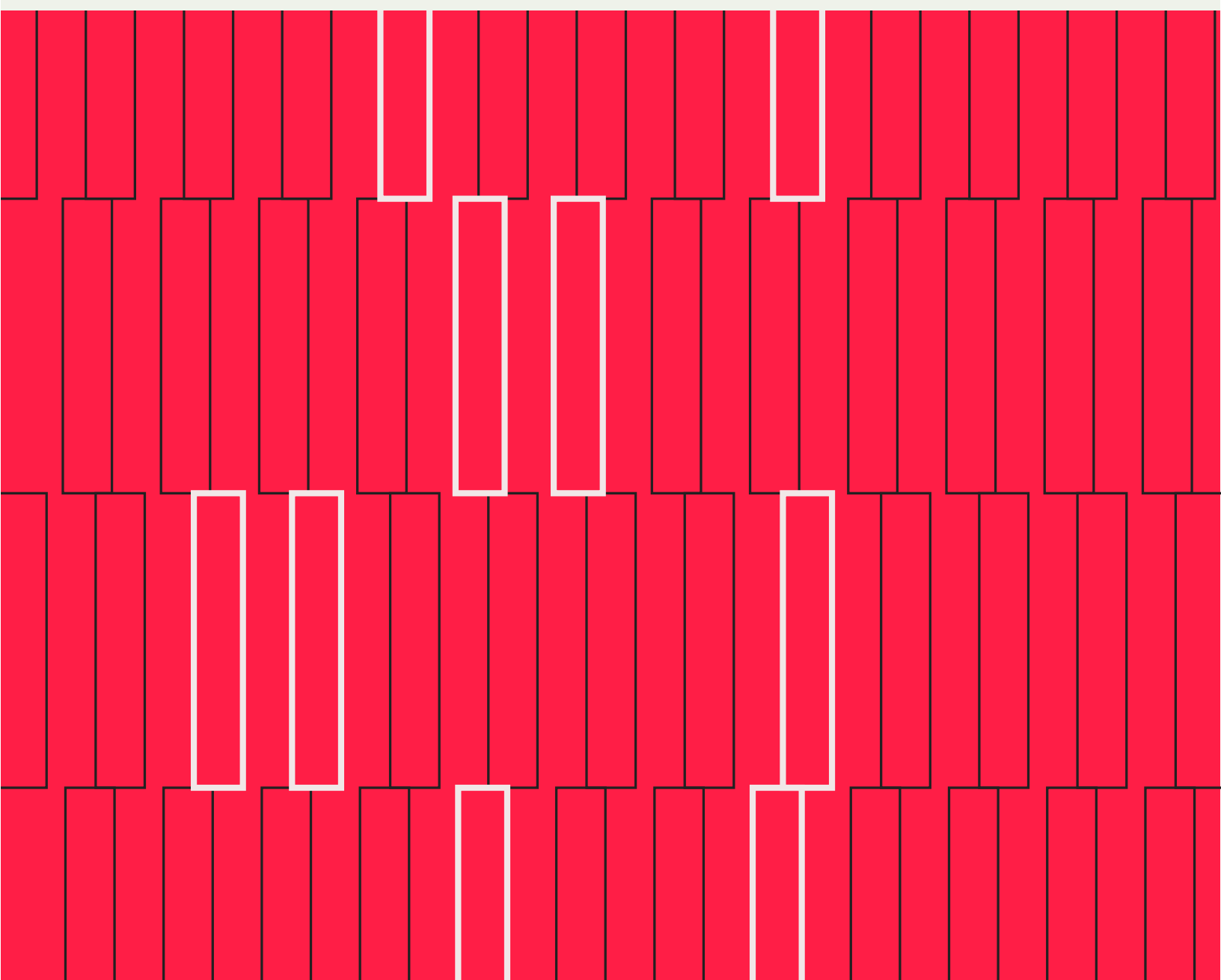


Table of Contents

03	Synthetic Fraud Overview
03	Product Overview
03	What the scores target
05	What the API returns
06	How to interpret the scores
06	How to use the scores
08	How we build the model
08	Raw data assets
09	Feature types
10	Analyst labels
10	Model metrics
10	Model development, management, and governance
11	Conclusion

Synthetic Fraud Overview

Synthetic identity fraud is one of the fastest growing types of financial crime in the U.S. It often goes undetected and costs financial institutions several billions dollars a year in losses. With the right solutions and strategies in place, it is now possible for a financial institution to accurately identify synthetic identities coming in the front door, at the point of application, before they infect a balance sheet.

SentiLink defines synthetic fraud as a name, DOB, and SSN combination that does not correspond to an actual, cohesive person. There are two types of synthetic fraud: first-party and third-party. SentiLink classifies a synthetic identity as “first-party” when a real person uses their true name and DOB but an SSN that does not belong to them. First-party tactics allow fraudsters to hide damaging information about their credit histories while passing standard KYC checks with their real government-issued ID. A “third-party” synthetic, however, is a wholly fabricated combination of name, DOB, and SSN, with no link to an actual person.

Product Overview

SentiLink’s Synthetic Score API returns a score indicating the likelihood that the identity submitted in an application is synthetic. Scores range from 0 to 1000, with higher being more risky. Organizations, including financial institutions, fintechs, cryptos, and non-bank lenders, ingest this model at the point of account opening through a real-time, hosted API.

What the scores target

As described, SentiLink’s Synthetic Score (“the Score” or “the Synthetic Score”) targets applications with a combination of name, DOB, and SSN that do not correspond to a single cohesive person. However, there are gradations of how severe these can be, ranging from a non-malicious typo or an immigrant using an ITIN because they don’t have an SSN to a fraudster using an identity they’ve fabricated to borrow money with no intention of repayment.

Our model is thoughtful about which of these we target. We have a complete taxonomy of these definitions, which include:

first_party_synthetic: name and DOB belong to the applicant who has an issued SSN/ITIN, but the input SSN/ITIN is non trivially different from the true SSN/ITIN (not a typo).

third_party_synthetic: name, DOB, and SSN do not belong to a single individual; none of the name, DOB, and SSN have an affiliation with the individual who is attempting the application.

idt_synthetic: the fraudster is using a real base identity and inserting an SSN that doesn't belong.

benign_synthetic: the applicant is using an SSN which does not belong to them, but this is because they do not have an SSN of their own.

friendly_fraud_synthetic: the applicant is using their own name and DOB but using the SSN of a relative.

test_application: the application information does not correspond to a consumer identity and consists of test data generated by the company.

pedantic: non-malicious typo in the SSN field.

malicious_pedantic: malicious and deliberate fiddling with SSN/ITIN (contrast with the pedantic label, which is an innocent typo).

clear: the person submitting the application is using their own identity.

The SentiLink Synthetic Score targets **first_party_synthetic**, **third_party_synthetic** and all malicious synthetic including fraud **friendly_fraud_synthetic**, but not the less severe variants of close family members sharing information without malicious intent.

What the API returns

In addition to the Score, the API also returns the three key explanatory model feature codes. These are useful for understanding what the model saw as most important for making its recommendation. Unlike “decline reasons” for FICO or Vantage that only explain the negative, these three codes provide context in both directions (i.e., why SentiLink believes the application is more fraudulent or less fraudulent).

Further, these explanatory codes are for internal use only and should not be shown to consumers in an adverse action notice or similar. In addition, we do not recommend using these particular reasons for constructing decisioning rules because the presence of a reason is only an indication that it impacted the model significantly, not that the application had a particular characteristic. For instance, the absence of R004 (“whether the supplied SSN aligns with the consumer’s DOB”) does not mean that the supplied SSN aligned with the applicant’s DOB, but rather that this characteristic of the SSN was not one of the three most significant contributors to the score. We have a separate attributes-based product that we recommend for use in constructing rules like this.

Each of SentiLink’s models is versioned, and the API also returns the model version used to score the request. For example:

```
{
  "transaction_id": "01GX7FHP-
N8AX-SQJKCAGM",
  "application_id": "APP-123456",
  "customer_id":
"ACCOUNTEXTTERNALNAME",
  "environment": "SANDBOX",
  "timestamp": "2023-04-10T22:52:5
7.980196465Z",
  "latency_ms": 100,
  "senticlink_synthetic_score": {
    "version": "1.8.1",
    "score": 175,
    "reason_codes": [
      {
        "code": "R016",
        "rank": 1,
        "direction": "more_
fraudulent",
        "explanation":
"Application cluster activity in
SenticLink consortium data"
      },
      {
        "code": "R010",
        "rank": 2,
        "direction": "more_
fraudulent",
        "explanation": "The
depth of the consumer's history with
this information"
      },
      {
        "code": "R004",
        "rank": 3,
        "direction": "more_
fraudulent",
        "explanation":
"Whether the supplied SSN aligns
with the consumer's DOB"
      }
    ]
  }
}
```

The SentiLink API docs contain the definitive list of all of the reason codes.

How to interpret the scores

The Synthetic Score ranges from 0 to 1000, with a higher score indicating a higher likelihood of fraud. Note that these scores themselves are not probability estimates of the likelihood of fraud, i.e. a score of 300 does not mean there is a 30% chance of fraud. The actual fraud rates by score vary between populations, depending on things like business line, applicant mix, and existing fraud controls. Still, across all of our partners, the approximate fraud rates by score band are:

SentiLink Synthetic Score Band	Fraud Rate
950-1000	100.00%
900-949	100.00%
850-899	98.38%
800-849	90.99%
750-799	76.63%
700-749	40.92%
650-699	32.42%
600-649	12.61%
550-599	10.00%
500-549	1.33%
450-499	0.00%
400-449	0%
0-399	0%

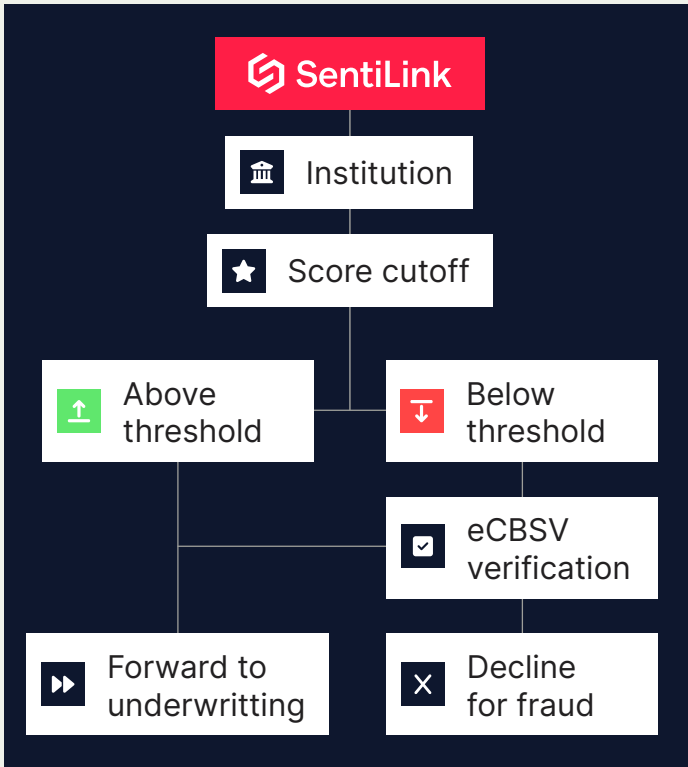
One unique challenge posed by first-party synthetic fraud is that traditional treatment strategies do not apply. For example, the individual attempting this type of fraud will:

- ✉ have a government-issued ID with their name and DOB
- ☎ provide a phone that's been tied to their name
- ✉ provide an email that's been tied to their name as well
- ? likely know all of the KBA type questions that could be generated from their fictitious identity since they created the identity]

How to use the scores

Score implementation

We recommend integrating the Score into your account opening flow and flagging applications above a specific cutoff (700 is a commonly used high-precision threshold, e.g.). All flagged applications should then be processed with an appropriate treatment strategy. For synthetic fraud (both first and third party), we recommend eCBSV as an appropriate treatment strategy.



We recommend using the Score in conjunction with eCBSV to validate risky applications. Ultimately, the Score cutoff selection and optimal treatment strategy implementation will depend on a number of partner-specific factors, including indicator precisions, economic costs of false positives and false negatives, manual review capacity, technology constraints, and desired consumer experience. SentiLink can help your team select the right threshold for you.

Score versions

Given the importance of stability and reliability for our partners, we do not change model versions underneath our partners without telling them. Instead, whenever we release a new model, we reach out to our

Why an eCBSV-only implementation is not advisable

The eCBSV verification service was created with the purpose of eliminating the risk of synthetic fraud by relying on the SSA, as a source of truth, to confirm that the provided information all belongs to a real person. Therefore, a “No Match” from eCBSV would ideally represent a fraudulent or otherwise suspicious application. This is not always the case, though. A notable shortcoming of eCBSV is the relatively high mismatch rate--i.e., when some aspect of the name/DOB/SSN combination submitted by a financial institution does not correspond to a record on file with SSA, and the system returns a “No Match.”

As of this writing the mismatch rate is around 8%, according to SSA. While this is trending slightly down from historical average monthly mismatch rates, it remains notably high for most financial institutions. We’ve looked closely at many of these mismatches, and when we are able to isolate a “No Match” response from the SSA that we can confirm is not fraud, oftentimes we can show that a mismatch on its own does not represent a credit risk, either. You can read more about it [here](#).

Given the substantial eCBSV mismatch rate, it is more effective if used as a treatment tool rather than a detection tool.

our partners to ask if they're interested in switching to it, while continuing to maintain older versions.

A partner may choose to run two versions of the same model side-by-side so that they can be compared. This is helpful in cases where a partner wishes to gain comfort with a new version before switching to that version for their internal decisioning. We also have the ability to score historical transactions under the new model so our partners can compare swap-in/swap-outs.

How we build the model

At a high level, SentiLink receives data from its partners and third-party sources, normalizes this data, derives features from the normalized data, and feeds this data into a statistical model to generate predictions. We test extensively during model building and deployment in order to ensure model quality.

Raw data assets




SentiLink's data assets fall into two categories: application data from partners and licensed data from third-parties.


Application Data: SentiLink receives applications for account opening from our partners via API. Application data consists of PII (name, date of birth, social security number, address, and optionally phone, email and IP address) along with timestamps for the applications' creation. This data is used to derive anonymized features across our consortium of partners, both in terms of clustered activity and velocity features. Here, "clusters" refers to deterministic cluster logic based on matching attributes. While SentiLink can improve its models using information aggregated from its network of partners, we never share specific PII originating from one partner to another. For instance, we might be able to see that a single IP address has been tied to a suspiciously large number of unique identities across our partner base over a short time, but we will not share what those unique identities are.

Licensed Data: SentiLink also acquires data from third-party vendors. This includes credit header information, which contains PII (name, date of birth, social security number), address history, consumer history length, and other information from credit reports stripped of tradelines and other FCRA-regulated credit information. SentiLink also licenses data, including phone records, phone carrier information, email records, bankruptcies, deceased data, IP information, and other public records.

Feature types

Internally, we group model features into different logical categories for feature versioning and development. The following are the primary categories for our Synthetic Score models:

-  **PII features:** features developed primarily on the PII from the application. Examples of features might be whether the application SSN is a randomly issued SSN or whether the SSN issuance aligns with the consumer's other information (e.g. an SSN issued before one's date of birth).
-  **Cluster features:** features developed based on clustered applications across SentiLink's partners. Examples would be noting that another consumer has applied using this SSN, or that it appears that this consumer has applied to another institution using a different SSN.
-  **Manifest features:** features based on the Manifest data, which consists of SentiLink's view of the distinct consumers in the United States derived from the credit header file. Examples of features would be the length of history that the consumer identity has existed, whether there are ties to addresses with known fraudulent activity, or whether the consumer has longer history with a better SSN.

-  **Contact information features:** features with additional focus on ownership and activity of specific points of contact: phone, email, address, IP address.

We are careful when building our models or scoring applications to compute our features only using data that would have existed as of the time of the application. For example, suppose the application occurs on May 1, we will only use data that we had before May 1, even if there is important data we gathered after May 1 that would be relevant to deciding whether the application is fraudulent. We refer to this as "as-of" feature calculation, which ensures consistency between scores returned in retrostudies and our production API.

Analyst labels

We use our fraud analysts’ manual labels as the source of truth by which we train the Synthetic Score model. These labels are the dependent variables for the model estimators. As described above, we maintain an internal taxonomy of labels, such as **clear**, **first_party_synthetic**, or **idt_synthetic**. SentiLink regularly audits past analyst labels, to both check their general accuracy as well as to confirm that they conform to our most recent fraud taxonomy.

Importantly, we do not use or require partner-contributed labels or performance data to train the Synthetic Score model. Labeling practices and taxonomies vary across partners, depending on things like industry, business line, or applicant mix. Using only internally-created labels allows us to maintain consistency and generalize better across industries and use cases. Moreover, non-use of performance data allows the Score model to work in settings without clear performance data, such as with applications for checking accounts. We do, however, use partner labels to QA our own labels.

Model metrics

When building models, we use various modeling metrics while tuning our features, parameters, and modeling methodology. Below are estimates of model performance on live partners using our analyst labels as the source of ground truth.

Metric	Mean across partners
F1	0.895
AUC	0.991
KS	0.940

Please note these are not business metrics, and depend as heavily on our population mix (more “obvious” fraud in the population inflates these numbers) as they do on modeling quality. To assess the quality of our models for your population, we encourage you to compute business metrics such as “total amount of dollars lost among applications flagged” or “number of new applicants who can be verified” rather than modeling metrics.

Model development, management, and governance

As part of model development and model risk management, SentiLink maintains extensive validation, stress testing, sensitivity analysis, quality assurance, and documentation processes.





Model testing and validation: SentiLink does extensive testing while building models. This includes evaluating models on different testing sets, including fully random samples, examining variable partial plots, and conducting swap-in/swap-out analysis compared to prior models. Importantly,

all model evaluation is done with out-of-sample results in order to most closely mirror the results we would return in production without “cheating.”



In particular, when performing cross-validation or out-of-sample testing, we group similar applications together. A fraudulent identity will generally have multiple loan applications in our dataset, and a subtle form of cheating can occur if one such application is in the training set and another application with nearly the same information appears in the test or validation set.

Significant additional detail on this process is available in our model governance document, described below.

Quality assurance: SentiLink’s extensive QA process for productionalizing models includes, but is not limited to, the following components:

-  **Data normalization and QA:** we normalize and QA data to ensure high-quality inputs to our models.
-  **Code review:** all changes have to go through a code review process and receive approval from another technical employee before being merged.
-  **Tests:** we use unit and integration tests to ensure our code exhibits its expected behavior.
-  **Live tracking:** we log information about API calls as they process through our system, logging warnings and errors if certain conditions occur that are

unexpected. Warnings and errors are assigned to technical staff and responded to immediately.

-  **Analyst-data scientist feedback loop:** SentiLink recognizes the importance of human feedback to maintain the quality of statistical models. As such, we maintain communication between the data science and fraud analyst teams at numerous points in the process.
-  **Score distributions:** before rolling a new model out to production, we will check score distributions subsetted to specific populations, to ensure that scores correspond to our expectations.

Model Risk Management: SentiLink works closely with model risk management teams within many of the largest financial institutions in the United States to help them better understand our model development and risk management processes. As such, we maintain clear and consistent documentation of our internal processes, including a 40-page model governance document that goes into more detail on all of our model building processes and controls. In addition, our model undergoes annual reviews by a third-party auditor, who assesses it for fair lending issues. Both our model governance document and our third-party reports are available upon request.