Prove

Identity Verify

Model Governance & Documentation for CIP Compliance (USA)

August 2022

Prove

Table of Contents

Executive Summary

This document contains technical and business documentation on Prove's Identity Verify API ("Identity Verify"). In its current version, the API performs identity matching and verification using a variety of authoritative data sources procured by or in partnerships with Prove. This document further outlines the details and scope of the API's match score models and other match elements: their development, performance, validation, assumptions, and limitations. The documentation also explains how Identity Verify can be used to help organizations comply with Customer Identification Program (CIP) regulations.

Business Purpose and Uses

In an increasingly digital economy accelerated by COVID-19, an increasing focus on fraud prevention is essential for many Prove customers. The purpose of Identity Verify is to enable a Prove customer (e.g., bank or other financial institution, insurance company, retail company, etc.) to verify an individual consumer's identity using a phone-first approach that a) validates consumer identity attributes as well as b) confirms ownership of their mobile, landline, and VOIP phone number ("Phone Number"). Additionally, validating a consumer's identity through their Phone Number enhances the consumer onboarding experience, ultimately reducing friction, improving customer service, and implementing top-tier data protection and security controls.

Verification goes beyond just matching name and address to the Phone Number. In most cases, it is capable of matching Social Security Number (SSN) or the last 4 digits of SSN, Date of Birth (DOB), driver's license or state ID information (i.e., number and state), and email address. Prove customers such as banks or other financial institutions, subject to such customers' internal legal and/or compliance team oversight, may choose to use the output of the results to comply with Customer Identification

Program (CIP) regulatory requirements (also known as “Know Your Customer” or “KYC” requirements).

Use Cases

1. Customer onboarding and/or account opening
2. Customer servicing (e.g., logged-in customer adds new Phone Number to their profile)

Response Interpretation

Identity Verify returns a response with a “verified” field of “true” or “false”, meaning the Phone Number is verified (or not) as owned by the person whose information is submitted. When the “verified” field is “true”, the identity attribute matching results can also be used to indicate whether each submitted identity attribute matches the person.

Model Development

Prove's proprietary logic leverages phone association and other information across our authoritative data sources to find the appropriate identity that matches to the customer input data. For each transaction associated to a specific Phone Number, Prove generates a verification response (e.g., "true" or "false") and field-level matching scores (e.g., match score for the "address" field is from 0-100) to verify the Phone Number ownership. The field level matching leverages input attributes to match the identity for which the response output (verification response and matching scores) can be used to decision.

Model Inputs

1. Phone Number (MSIDSN)

2. Name (at least one)
    - First
    - Last

3. Address (all fields preferred)
    - Street address number

- Street
- City
- State/Region
- Postal/zip code

4. SSN or last 4 digits (optional)

5. Date of Birth (optional)

6. Email Address (optional)

7. Driver's license or State ID (optional)
    - Number
    - State

**Note*: Inputs #4, #5, #6, #7 are matched to Prove's authoritative source data, but Prove does not currently use the field level matching score algorithm (so only Boolean True/False is returned).

Data Sources

Prove

Prove uses carrier and authoritative non-carrier data sources to compile phone and identity intelligence and generate the verification output results for Identity Verify.

Our data sources and partners include national credit bureaus as well as companies in the “core telecom” infrastructure, service provider billing systems, banking, public records, utilities and prepaid processors (for the 40% of prepaid phone lines that are funded with debit/credit cards). The “core telecom” infrastructure includes over 3,000 phone companies in North America (i.e., service providers).

When performing identity verification in instances where Gramm-Leach-Bliley Act (GLBA), or similar regulated sources are used, credit bureau data often meets a customer bank’s requirements [e.g., to satisfy a bank’s Customer Identification Program (CIP)].

Model Outputs

Prove's authoritative data sources are queried to locate the identity associated to the input Phone Number, which is then matched to the customer input data.

- If no customer input data is submitted, the "verified" field returns as "false", and no attempt to match phone ownership is performed.
- If customer input data is submitted, then an attempt to match phone ownership is performed.

More specifically, the parameter "details = true" must be passed in the API request and *at least* either the "name" field or the "SSN"/"last4" field as input data for there to be an attempt to verify phone ownership and provide identity attribute match results in Prove's response.

Matching Logic

Given that fraud attack methods are becoming increasingly complex due to increased intelligence and technical capabilities, Prove understands the importance of performing analysis beyond exact matching to inform its proprietary algorithms and scoring.

The following logic is included in the “name” match model for fields submitted in each request:

- Nickname matching
- First name/last name transposition
- Name contained within
- Accommodation for name matching for those with a changed last name due to marriage
- Mother and father’s name concatenated or split within middle and last name
- Levenstein distance calculation (i.e., fuzzy match logic)

The following logic is included in the “address” match model for fields submitted in the request:

- Match Scoring (0-100) for the Street Address (house) Number
- Boolean matches:
    - Street
    - City
    - State
    - ZIP3
    - ZIP5

    - ZIP9
- Distance (miles) – Calculated between the normalized input address and the address that is being matched to “as the crow flies”.
- Full Address match score (0-100) - designed to be similar to human discernibility (to the eye, do these addresses look the same) and allows for small typos. Since postal code is somewhat more specific than city/state, we look for a match in postal code and if we find one, we don’t need to look at city/state, but if there is not a postal code match, we can make up for that with correct city/state. The street name and house number are compared using fuzzy string matching, Damerau–Levenshtein for the house number and Jaro-Winkler for the street name; these fields are where we allow for small typos. Scores for all these components are compiled for a full address match score. A score of 100 is considered to be a matching score that balances both Type I and Type II errors.

The rest of the Identity Verify parameters listed below use Boolean matches for any fields submitted in the request:

- Date of Birth
- SSN or last 4 digits (if both are submitted, matching is done with full)
- Email Address

- Driver's License or State ID (e.g., number and state)

Score Component Framework

The method(s) used to calculate the name and address scores uses the logic described above and checks it against the information submitted by customers through the Identity Verify API. Prove *only* matches client input data and does not supplement customer input data as inputs with any other data or data sources.

The match results are then run through weighted algorithms; the weight applied is influenced by how much variance was found between customer input data and Prove queried data (e.g., which fields returned true/false, etc.). These calculations then return a "verified" status (e.g., "true" or "false") with a total score for the name matching and, where applicable, a total field level match score for the address matching. Each field level match score falls anywhere between 0–100. As was mentioned prior, these scores contribute to the verification of the consumer with the following logic:

- ***Either*** a name score *greater than or equal to* 70 ***or*** address score *equal to* 100 contributes to “verified” returning “true”.
- ***Both*** name score being *less than* 70 **and** address scores being *equal to* 100 contributes to “verified” returning “false”.

**The value of 70 and 100 is configurable as well as the combination logic (e.g., NAME only, ADDRESS only, NAME or ADDRESS, NAME and ADDRESS)*.

Note that if SSN (or the last 4 digits of SSN) is submitted by the consumer, and neither name field is submitted, the name and address model is not enabled, and “verified” matching is performed using the SSN/last 4 model.

- If the SSN/last 4 submitted matches the data found, “verified” returns “true”.
- If the SSN/last 4 data does not match the submitted information, “verified” returns “false”.

If either name field is submitted, the SSN/last4 simply match to show “true” or ”false” for that SSN/last4 field only and has no bearing on the “verified” result.

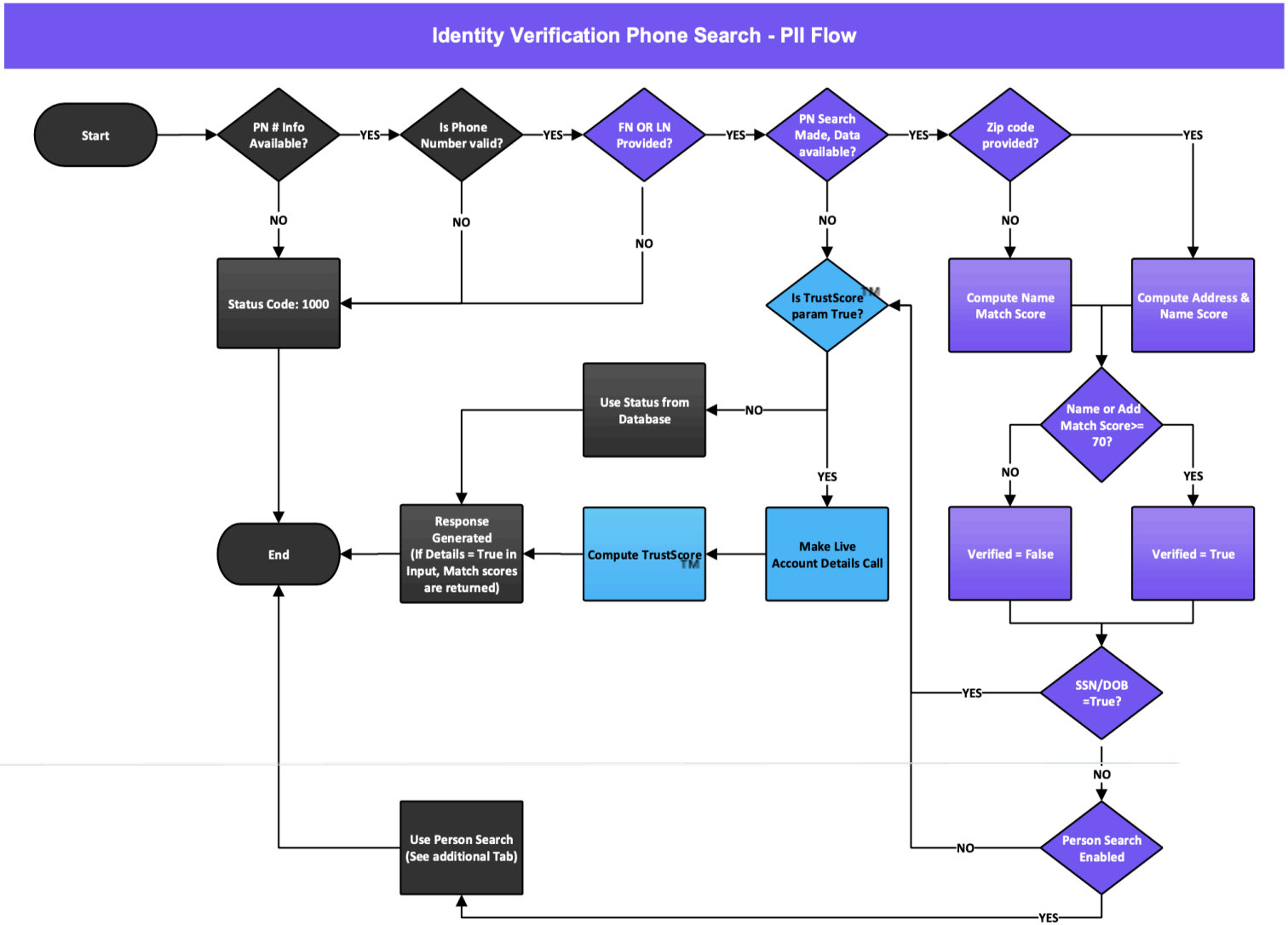
Additionally, a death indicator found present on the associated identity causes “verified” to return “false”. The rest of the parameters are simply informational, and do not contribute to the “verified” response.

Reason Codes

Along with the various scores described above, Prove provides specific reason code(s) in the API response that give further insight into why a certain field level score was returned and/or indicate one or more specific risk(s) associated with the identity. This is to assist our customers with making informed decisions in managing their respective risk. Prove API overview and reference documentation is available here which outlines the available reason codes, what they represent, how to interpret them, and how the interpretation might be applied.

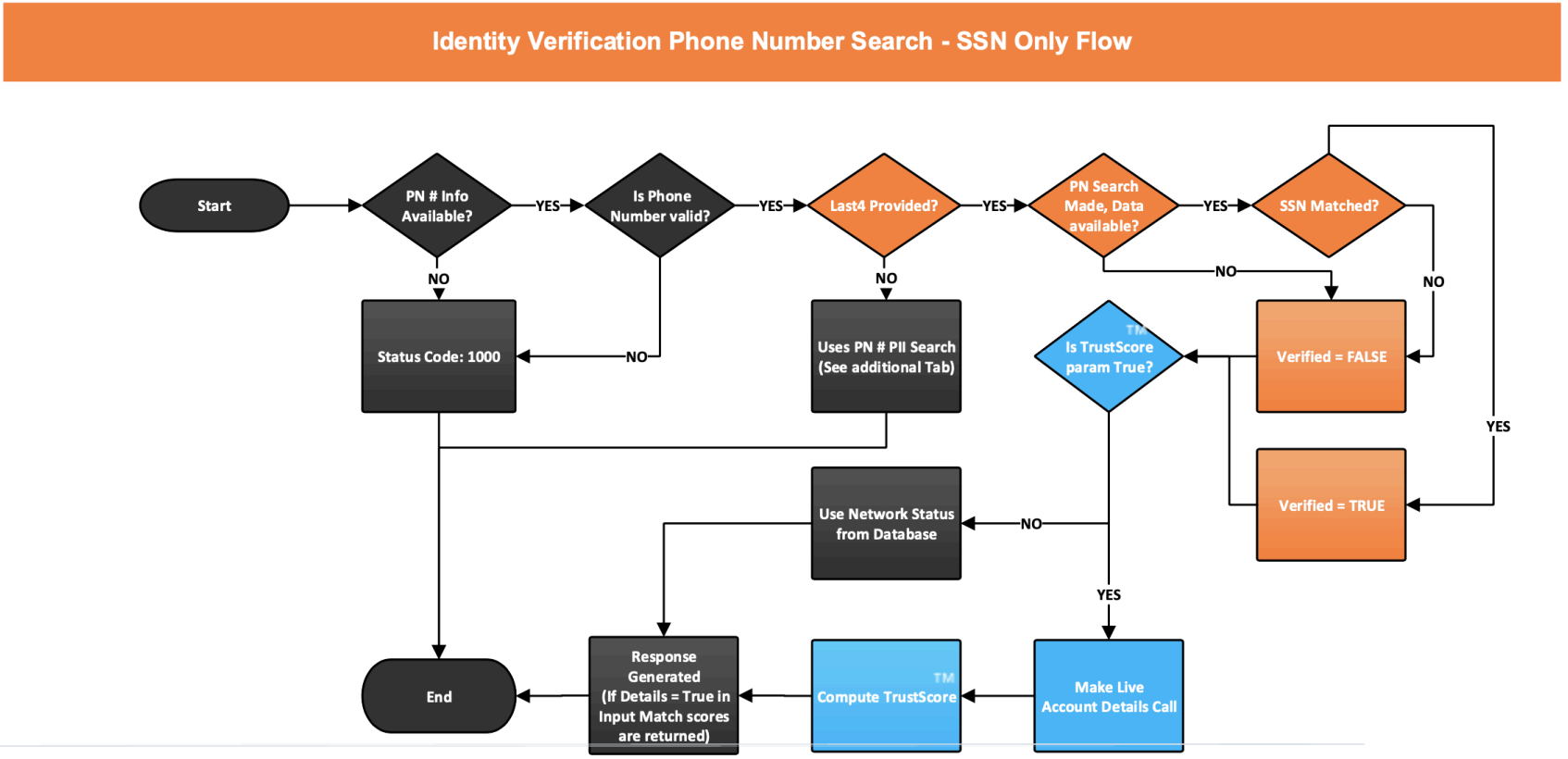
Prove

Full Model Structure



Prove

Full Model Structure

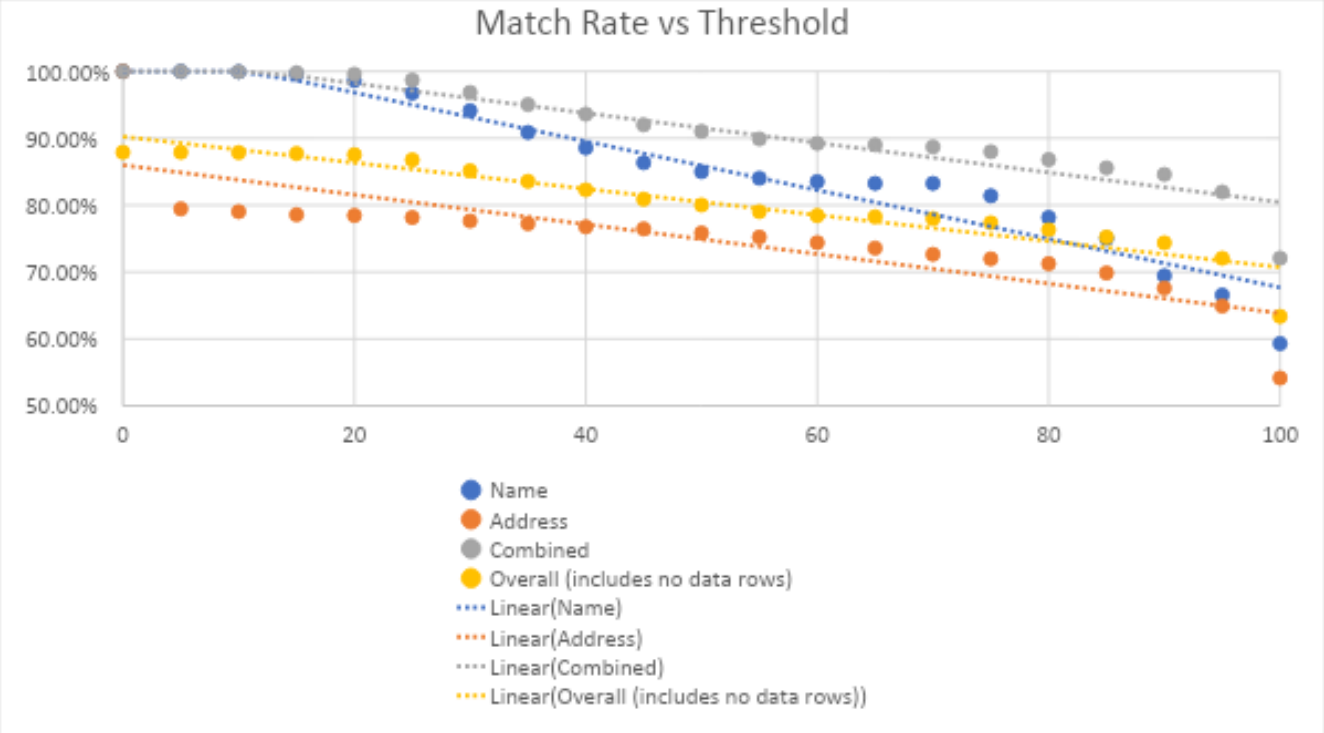


Performance and Ongoing Validation

The performance of the matching system was tested using a file of known-good address and user data comprising 2,271 unique records. Using this data, the file was compared to our data sources for matching rate and the results were analyzed to determine the level of matching that provided the highest match rate without creating false positives or eliminating valid matches.

Given that only approximately 90% of people have accurate records with the third-party groups, the best-case would be a 90% match rate.
It was found that using a score of 70 found in either name or address produced an approximately 87% positive match rate to the data. Going below a match score of 65 produced clearly inaccurate matches during analysis, while above 70, several correct matches were ruled out.

The attached graph shows the rate at which matches were found at each threshold.



Scope and Limitations

Identity Verify is designed to be used as security or fraud control, and *not* for credit approval decisions, which typically are more subject to variable exclusions due to illegal discrimination in lending. It is important to note that Identity Verify does not take any of the factors considered as prohibitive basis under fair lending regulations, such as sex, religion, race, color, sexual orientation, etc. Identity Verify name and address matching models, along with the rest of the API's matching parameters, focus on quantifiable matching attributes to determine verification of phone and account ownership in transactions.

Assumptions

Bad actors behave in a certain manner and take advantage of changes during the life cycle of a Phone Number, as well as other personal information. The Identity Verify ownership verification can help mitigate the risk associated to these types of scenarios by providing a rich set of output signals to help inform customer decisioning.

Limitations

The query parameter of details (e.g., customer input data) must be submitted as true in order to return detailed field scoring, though matching is still performed when submitted as false. As discussed in various sections above, the model and matching logic relies on valid information submitted by our customers to match with Prove's queried data, so only data elements passed to Prove that meet the query parameters can be included in the model response. For example, if no address is submitted, or certain fields of the address are omitted, that information cannot be matched, and Prove will not return a response for that specific data element (e.g., address).