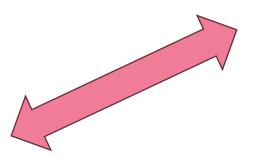


# **Pipeline**

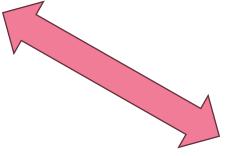
Extracted Data



Chloé Anaïs Nicolas and the RM crossed paths at a tech conference in Paris. They attended several keynote sessions together and realized they had a shared vision for the future of technology investments. Chloé Anaïs Nicolas is a seasoned professional with over 40 years of experience. Seeking a trusted partner for wealth management, she joined Julius Baer for its comprehensive and tailored services.



#### Compare



RÉPUBLIQUE FRANÇAISE / French Republic

PASSPORT Code PASSPORT No. PASSEPORT FRA WV1305067

Surname Given Name(s)
NICOLAS CHLOÉ ANAÏS

Birth Date Citizenship

21-Feb-1961 French/FRANÇAISE

Sex Issue Date

19-Mar-2018

Expiry Date Signature 18-Mar-2028 Chloé Nicolas

P<FRANICOLAS<<CHLOÉ<ANAÏS<<<<<<<<<<<>W1305067FRA610221<<<<<<<



----- Client Information -----

Name:

Chloé Anaïs Nicolas

Address: Rue Cujas 10

36022 Dijon

Country of Domicile:

France

Date of birth: 1961-02-21

Nationality: French

### Jaro-Winkler Distance

- To compare different occurences for the same information type we use the Jaro-Winkler distance.
- It measures edit distance between strings.
- To compare the data in tabular form with the full texts we use **LLM** based **feature extraction**.

## Feature Engineering

Additional hand-engineered features were added to account for

- Mismatches in passport MRZ
- Outdated passports
- Missing salary information
- Incorrect country codes



### Model

- These features are then used with a LightGBM model
- Hyperparameters were tuned using cross-validation
- We achieved 0.95 accuracy on a hold-out validation set, with most errors being False Negatives.