

# SME Insight Navigator

A comprehensive dashboard and AI agent for modern risk assessment

# The Path to Smarter SME Evaluation

**01**

## The strategic challenge

Italian SMEs represent the economy's backbone but suffer from sparse structured data, making risk assessment difficult for financial institutions.

**02**

## The SME insights extractor

Our AI User Agent automates the collection of sector-specific signals from an allowed list of public sources.

**03**

## The opportunity

Digital footprints offer a "relatively reliable foundation" for deeper analysis.

**04**

## The SME dashboard creator

These signals are synthesized into our interactive Dashboard.



# The SME insights extractor

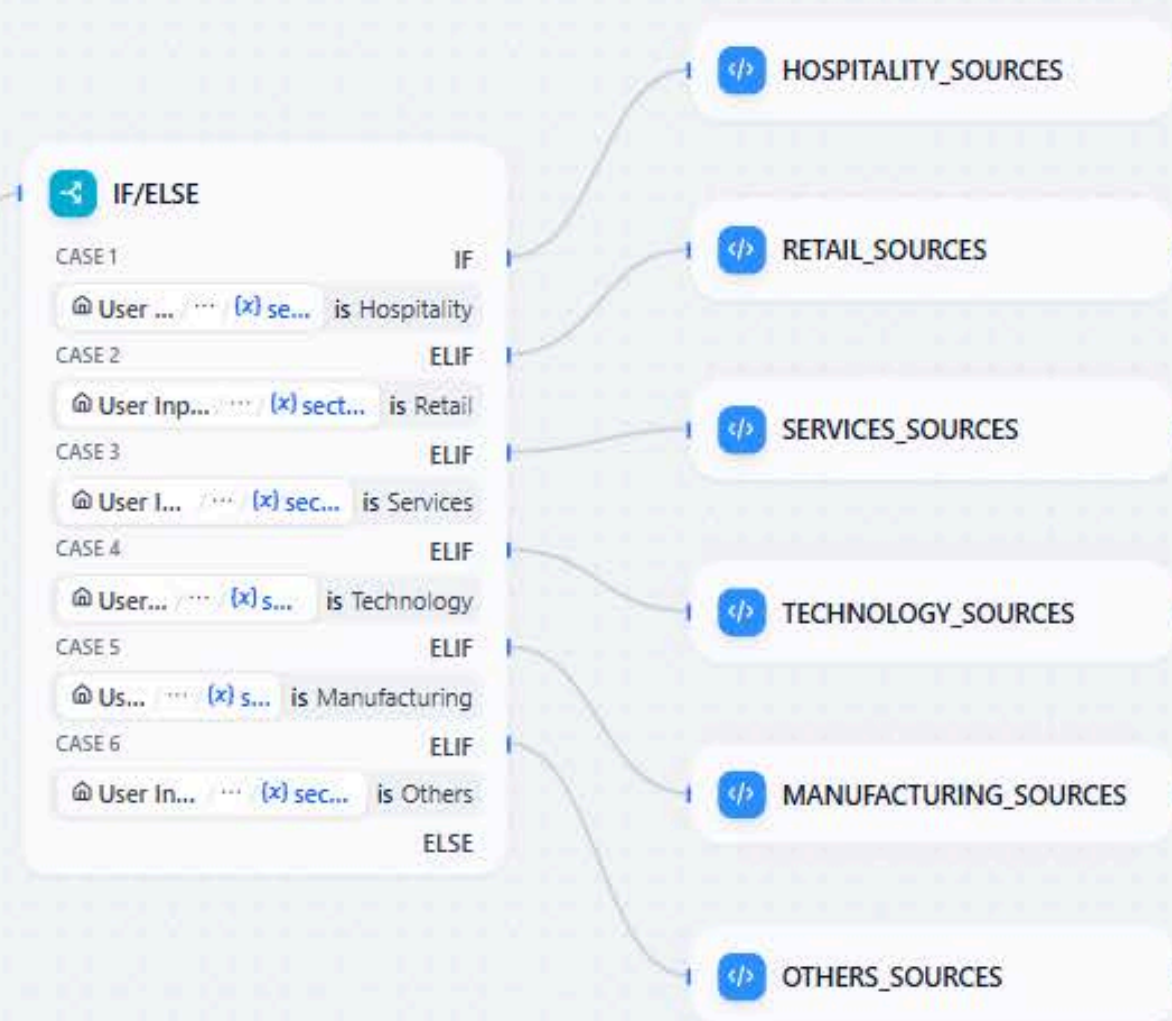
## Company information input

START

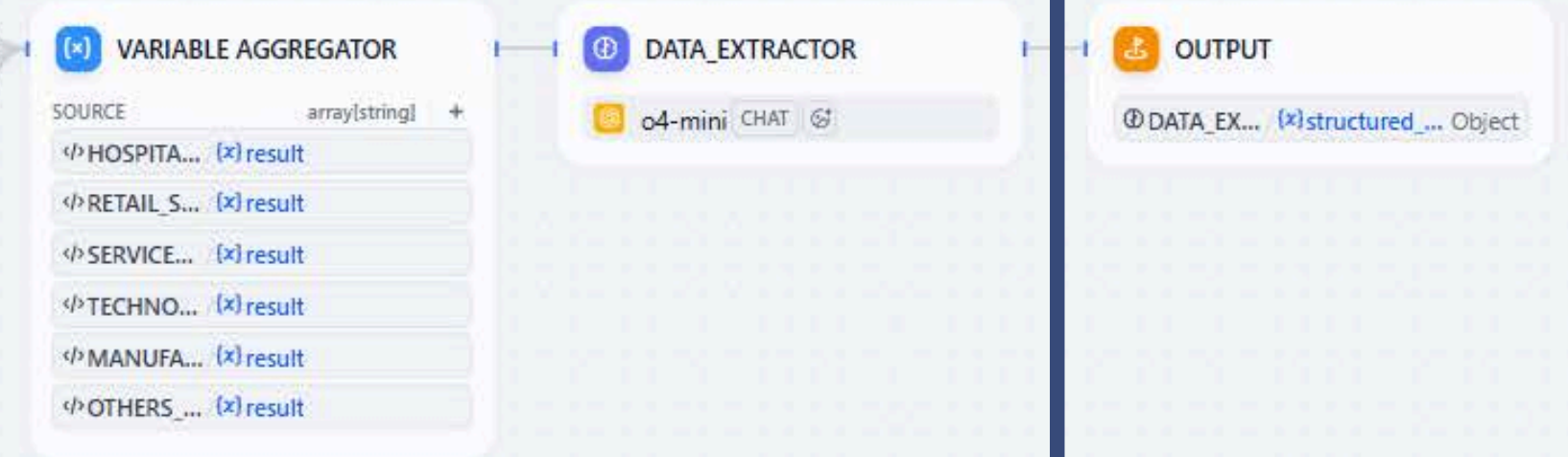
**USER INPUT**

- [x] company\_name REQUIRED
- [x] sector REQUIRED
- [x] website\_url REQUIRED
- [x] city
- [x] VAT\_number

## Choosing the data sources set



## Data extraction



## JSON file output

# Mapping Alternative Data Sources

Sector-specific

Social media set

Booking aggregators set

Google sources set

Financial info set

`</>` HOSPITALITY\_SOURCES

PYTHON3  $\diamond$

```
1 def main() -> dict:
2     sources = {
3         "social_medias": [
4             "https://instagram.com",
5             "https://facebook.com",
6             "https://linkedin.com",
7             "https://x.com",
8         ],
9         "booking_aggregators": [
10            "https://www.tripadvisor.com",
11            "https://www.airbnb.com",
12            "https://www.booking.com",
13            "https://www.trivago.com",
14            "https://www.paginegialle.it",
15            "https://www.paginebianche.it",
16        ],
17        "google": [
18            "https://maps.google.com",
19            "https://trends.google.com/trends/",
20        ],
21        "financial": [
22            "https://www.registroimprese.it",
23            "https://www.registroimprese.it/ricercaext",
24            "https://www.agenziadelleentrate.it",
25            "https://www.bancaditalia.it/servizi-cittadino",
26            "https://italianbusinessregister.it/en/",
27        ],
28    }
```

# The Data Extractor

**01**

## What is it?

LLM-powered agent that:

- Collects public business signals data
- Only sector-specific sources are allowed for search

**02**

## Data Categorization

Categorizes data in:

- Digital presence & engagement
- Customer perception (ratings, reviews)
- Operational indicators (hours, delivery)

**03**

## Privacy & Compliance

Hard constraints:

- No personal data collection
- No bypassing logins, paywalls
- Blocked sources → status flag

**04**

## JSON structured output

Divided in 3 metadata sections:

- Entity (company, match confidence)
- Run (timestamp, sources used/blocked)
- Records (source-specific extracted fields)

# Future Roadmap

## Strategic Model Evolution

### ESG Integrity & Greenwashing

Criteria:

- *has\_verified\_iso14001*
- *greenwashing\_flag*

Impact: Protects banks from reputational risk by verifying **genuine sustainability** claims.

### Intellectual Capital

Criteria:

- *active\_patents\_uibm*
- *trademarks\_count*
- *university\_partnership*

Impact: Validates the SME's competitive "moat" and long-term **innovation potential**.

### Geospatial & Physical Risk

Criteria:

- *ispra\_flood\_risk\_zone*
- *seismic\_zone\_rating*
- *local\_crime\_index*

Impact: Ensures **collateral protection** by assessing the physical vulnerability of warehouses and storefronts.

# APIs & Sources

## ESG Integrity & Greenwashing

Data Point	Primary Source	Reliability
<i>has_iso14001</i>	<b>Accredia</b> (Italian Accreditation Database)	High (Official)
<i>greenwashing_flag</i>	<b>SME Website</b> (via NLP analysis of sustainability claims)	Medium (Predictive)

# APIs & Sources

## Intellectual Capital

Data Point	Primary Source	Reliability
<i>active_patents</i>	<b>UIBM</b> (Italian Patent and Trademark Office)	High (Official)
<i>uni_partnerships</i>	<b>NLP Analysis</b> of Press Releases & LinkedIn	Medium (Social Signal)

# APIs & Sources

## Geospatial & Physical Risk

Data Point	Primary Source	Reliability
<i>ispra_flood_risk</i>	<b>ISPRA</b> (Hydrogeological Risk Open Data)	High (Official)
<i>seismic_zone</i>	<b>INGV</b> (Seismic Hazard Maps)	High (Official)

**Thank  
You**

for your time  
and attention