ALBERTSCHOOL UniCredit

BDD5

SME Insight Navigator

A comprehensive dashboard and AI agent for modern risk assessment

Asriev Artem, Limonta Tommaso, Premazzi Carola, Radaelli Gregorio

Context & Why This Matters

01

The institutional reality

UniCredit and Italian banks assess SME creditworthiness primarily through balance sheet data filed with the Agenzia delle Entrate. By the time that data is evaluated, it is often 12–18 months old; a structural lag embedded in the annual filing cycle.

02

The gap traditional data leaves

Financial statements capture the past. They do not reflect operational vitality - whether a business is busy today, well-regarded by customers, or losing ground to competitors. Reputational health is invisible to the balance sheet.

03

The opportunity digital transformation created

Those same SMEs now leave rich digital footprints: reviews, social signals, registries, booking data. A decade ago this data was noise. Today, AI and NLP can transform it into structured, explainable credit signals.

The question we set out to answer

“What if a credit officer could see a business’s real-world performance signals before even opening the balance sheet?”

The Path to Smarter SME Evaluation

INPUT

Italian SMEs suffer from sparse structured data, making risk assessment difficult for financial institutions.

EXTRACTION

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

SCORING

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

OUTPUT

These signals are synthesized into our interactive Dashboard.

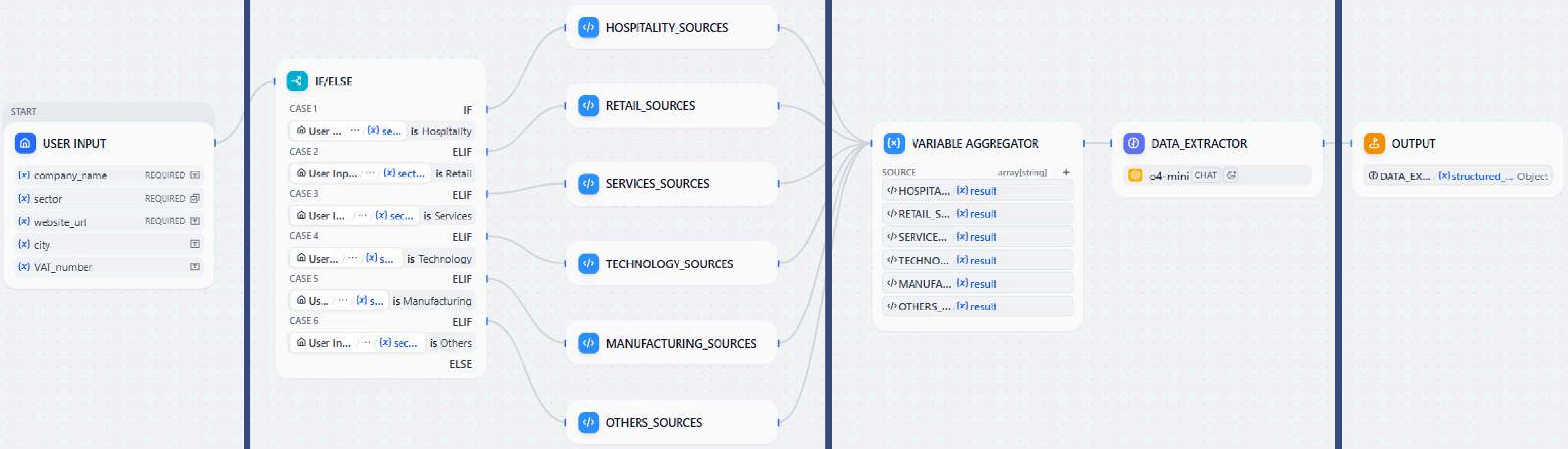
The SME insights extractor

Input: SME information

Choosing the data sources set

Data extraction

Output: JSON file



Mapping Alternative Data Sources

Sector-specific

Social media set

Booking aggregators set

Google sources set

Financial info set

HOSPITALITY_SOURCES

PYTHON3

```python
def main() -> dict:
    sources = {
        "social_medias": [
            "https://instagram.com",
            "https://facebook.com",
            "https://linkedin.com",
            "https://x.com",
        ],
        "booking_aggregators": [
            "https://www.tripadvisor.com",
            "https://www.airbnb.com",
            "https://www.booking.com",
            "https://www.trivago.com",
            "https://www.paginegialle.it",
            "https://www.paginebianche.it",
        ],
        "google": [
            "https://maps.google.com",
            "https://trends.google.com/trends/",
        ],
        "financial": [
            "https://www.registroimprese.it",
            "https://www.registroimprese.it/ricercaext",
            "https://www.agenziadelleentrate.it",
            "https://www.bancaditalia.it/servizi-cittadino",
            "https://italianbusinessregister.it/en/",
        ],
```

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)

Financial Data Integration

Agenzia delle Entrate

RETROSPECTIVE

- VAT number cross-validation
- Legal entity & incorporation status
- ATECO sector code classification
- Registered address verification
- Annual filing history

Limitation: annual filings describe past health. Data is typically 12-18 months old at time of assessment.

+

PSD2 / Open Banking

REAL-TIME

- Cash flow patterns & revenue cycles
- Payroll regularity & supplier payments
- Seasonal revenue fluctuations
- Real-time operational health signal
- Requires explicit SME consent (GDPR)

Bridge: real-time cash flow validates digital signals, closing the data freshness gap.

Scoring Methodology

Transparent, weighted formulas; configurable by the risk team

95 /100 Reputation & Customer Demand

50% Rating Quality (avg 4.59 / 5)

30% Review Volume (log x10, 8,236 total)

20% Sentiment Presence

79 /100 Digital Presence & Engagement

60% Social Reach (log x10, 276K followers)

30% Post Recency (last post Feb 14, 2026)

10% Website Presence

Actionable Results

Credit pre-screening flags, translating scores into credit officer decisions

GREEN - Expedited Review

Conditions:

- Reputation Score > 80
- Digital Presence Score > 70
- PSD2 cash flow consistency: stable
- Registry data: complete

Action:
Fast-track SME loan review. Reduced documentation burden.

AMBER - Standard Review

Conditions:

- Mixed signals: one score < 60
- Limited digital footprint
- PSD2 data: seasonal volatility
- Registry: minor gaps

Action:
Proceed with standard credit analysis. Request supplementary docs.

RED - Enhanced Due Diligence

Conditions:

- Reputation Score < 40
- No digital presence detected
- PSD2: irregular cash flows
- Registry: incomplete or missing

Action:
Flag for enhanced due diligence before proceeding.

Compliance & GDPR

How the system handles data responsibly

01 Public Data Only

All signals extracted from publicly accessible sources (business directories, registries, and public social profiles). No logins, no paywalls bypassed.

02 No Personal Data Collection

Individual reviewer names and personal identifiers are never stored. Business contact info is treated as commercial data, not personal data.

03 PSD2: Consent-Based Access

Open Banking data requires explicit, revocable SME consent under the PSD2 framework. Access is scoped, audited, and time-limited.

04 Full Audit Trail

Every data source attempted is logged with status, method, and timestamp. Blocked or unavailable sources are flagged: the system never fails silently.

05 Absence of Data is Not Penalised

SMEs with limited digital presence are not scored negatively. Missing signals are flagged as gaps, not deficits: they are recognised as coverage bias.

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
has_iso14001	**Accredia** (Italian Accreditation Database)	High (Official)
greenwashing_flag	**SME Website** (via NLP analysis of sustainability claims)	Medium (Predictive)

APIs & Sources

Intellectual Capital

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

APIs & Sources
Geospatial & Physical Risk

Data Point	Primary Source	Reliability
ispra_flood_risk	**ISPRA** (Hydrogeological Risk Open Data)	High (Official)
seismic_zone	**INGV** (Seismic Hazard Maps)	High (Official)

Conclusion

01

What we built

An alternative data extraction and scoring engine for Italian SMEs. It collects public digital signals and produces transparent, explainable scores.

02

What it is not

This system does not replace credit scoring models, balance sheet analysis, or the relationship banking judgment of a credit officer. It acts as a complementary pre-screening layer.