Fontys

Data Requirements

Phase 2

Poll,Laurenz L.P. van de
1-23-2024

Contents

Introduction 2
Data Types and Elements 3
Data Volume & Quality standards 4
Amount of data 4
Desired level of accuracy 4
Data Quality standards 4
Legal and Ethical aspects 5
Compliance with Data Regulations: 5
Privacy and Anonymization: 5
Ethical Use of Data: 5
Respect for Investor Autonomy: 5
Data dictionary 6
Location 6
Dictionary format 6
Essential information 6
Changes 6

Introduction

This document outlines the data requirements for the "Fundamental Analysis Rating" system, an innovative AI-driven project focused on ASR Nederland, a key entity in the insurance sector. Our aim is to harness AI to automate investment recommendations - "Buy," "Sell," or "Hold" - a process traditionally dependent on extensive expert analysis. This system is poised to revolutionize investment decision-making by offering data-driven insights, particularly within the insurance industry context.

The project's scope is exclusive to ASR Nederland, with an emphasis on the unique dynamics of the insurance sector. Although the project is academically oriented and constrained to a two-week period, it aims to demonstrate AI's potential in enriching investment analysis. Therefore, identifying and specifying the right data is crucial for the accuracy and relevance of our AI model's predictions.

The following sections will detail the specific types of data required, their sources, and the methodologies for their collection and processing. By clearly defining these data requirements, we intend to lay a solid foundation for developing a robust AI model that meets our objective of automating and enhancing investment advice for ASR Nederland.

Data Types and Elements

For the "Fundamental Analysis Rating" system, we will collect a variety of data types, each with specific attributes essential to achieve our project goals. Below, we outline the primary data types and list the corresponding data elements with their units and ranges.

- Numerical Data
  - Other Comprehensive Income (OCI)
    - Units: Currency (e.g. Euros)
    - Range: 0 - 1.000.000
  - Combined Ratio
    - Units: Percentage
    - Range: 0 - 200% (less than 100% indicates profitability)
    - Source: https://www.asrnl.com/investor-relations/financial-publications
  - Return on Equity (RoE)
    - Units: Percentage
    - Range: -100% - 100%
    - Source: https://www.asrnl.com/investor-relations/financial-publications
  - Price to Book ratio (P/B)
    - Units: Percentage
    - Range: 0 – 5
  - Outstanding shares
    - Units: Numbers
    - Range: 0 - ∞
    - Source: https://www.asrnl.com/investor-relations/financial-publications
- Time Series Data
  - Financial reporting
    - Time Interval: Annually
    - Range: Relevant from 2016, 10 June
- Categorical data
  - Analyst Consensus
    - Categories: Buy, Outperform, Hold, Underperform, Sell

Each of these data types and elements has been carefully selected to align with the project's goal of analysing ASR Nederland's performance and potential investment value. This will enable the AI system to perform a deep, data-driven analysis and provide valuable investment recommendations. The data will be sourced from company financial statements, market reports, and industry databases, ensuring a comprehensive and factual basis for the AI model's analysis.

Data Volume & Quality standards

This section addresses the crucial considerations of data volume and quality standards, which are pivotal for the predictive accuracy and reliability of our AI model in the "Fundamental Analysis Rating" system. Ensuring substantial, relevant, and high-quality data is fundamental to building an accurate and reliable model for financial analysis and investment recommendations.

Amount of data

- **Historical Depth**: For a comprehensive analysis, a minimum of 5 years of historical data for key financial metrics like OCI, Combined Ratio, RoE, P/B ratio, and Outstanding Shares is essential. This duration is chosen to capture a full business cycle and recent trends.
- **Periodic Updates**: The financial data, especially the Price to Book ratio and Outstanding Shares, should be updated at least quarterly to reflect the latest market conditions and company performance.
- **Analyst Consensus**: A continuous aggregation of the latest analyst ratings and consensus is necessary to keep the model updated with market sentiments.

Desired level of accuracy

- The model aims to match or exceed the accuracy level of current industry-standard financial analysis tools. This ensures that the stakeholders find the model's output useful for making informed investment decisions.
- Accuracy will be evaluated based on the model's ability to align with actual market performance and analyst ratings over time.

Data Quality standards

- **Source Integrity**: All data should be sourced from reputable and authoritative sources, such as ASR Nederland's official financial publications, recognized financial market data providers, and analyst reports.
- **Privacy Compliance**: Data collection and usage will adhere to relevant privacy laws and regulations, ensuring ethical handling of sensitive financial information.
- **Consistency and Completeness**: Efforts will be made to minimize missing or inconsistent data. Any gaps or anomalies in data should be documented with explanations, and where possible, rectified through data cleaning and validation processes.
- **Verifiability**: Data sources should be transparent and verifiable, allowing for cross-checking and validation of the data used in the model.

Legal and Ethical aspects

Compliance with Data Regulations:

- **Regulatory Adherence**: Ensure strict compliance with financial data handling regulations, such as the European Union's Markets in Financial Instruments Directive (MiFID II) and other relevant financial sector regulations. This includes adherence to standards for transparency and accuracy in financial reporting.
- **Use of Public Financial Information**: Understand and respect the legal nuances of using publicly available financial data of listed companies, ensuring that all data sourcing is legal and ethically justifiable.

Privacy and Anonymization:

- **Confidentiality of Sensitive Financial Information**: While most data for this project is publicly available, utmost care will be taken to protect any sensitive financial information that might be used. This includes internal performance metrics of ASR Nederland if utilized.
- **Data Anonymization and Security**: Ensure that any sensitive data, particularly if individual investor data is involved, is anonymized and secured to prevent unauthorized access or use.

Ethical Use of Data:

- **Bias and Fairness in AI Models**: Identify and mitigate potential biases in the AI model, ensuring that the investment recommendations are fair, unbiased, and based solely on financial data and trends, not influenced by external factors.
- **Transparency in AI Operations**: Maintain a high degree of transparency regarding the AI model's functionality, data sources, and the rationale behind its predictions. This should include clear communication about the model's capabilities and limitations.
- **Accountability for AI Decisions**: Establish clear lines of accountability for decisions made by or with the aid of the AI system. This includes a framework for reviewing and auditing AI-generated recommendations to ensure they align with ethical standards and investment best practices.

Respect for Investor Autonomy:

Informed Decision Making: Ensure that the AI system is used as a tool to aid, not replace, human decision-making in investments. The system should empower investors with insights while respecting their autonomy in making final investment decisions.

Stakeholder Engagement: Actively involve stakeholders, including investors and financial analysts, in understanding how AI is used in the project. This includes addressing any concerns about AI replacing human judgment in financial analysis.

By adhering to these legal and ethical standards, the "Fundamental Analysis Rating" system aims to be a reliable, transparent, and ethically sound AI tool in the landscape of financial investment analysis, particularly for evaluating ASR Nederland.

Data dictionary

Location

The latest version of the data dictionary is maintained within the current project notebook. This approach ensures easy access and reference during analysis and model development.

Dictionary format

The data dictionary is structured in a table format, providing key details for each data field, including the Dutch and English terms, descriptions, data types, range, units, source, and quality standards.

Essential information

Each entry in the dictionary includes:

- Dutch Term: Original data field name in Dutch.
- English Term: Translated data field name in English.
- Description: A brief explanation of what the data field represents.
- Data Type: Classification of the data (e.g., Numerical, Time Series, Text).
- Range: The possible values or extent of the data.
- Units: Measurement units for the data (if applicable).
- Source: The source from where the data is obtained.
- Quality Standards: The required standards for data accuracy and consistency.

Changes

All changes to the data dictionary should be documented here. For each change, include:

- Date of Change: The specific date(dd-mm-yyyy) when the change was made.
- Description of Change: A clear description of what was changed, added, or removed.
- Reason for Change: Brief explanation of why the change was necessary.
- Impacted Fields: List of data fields affected by the change.

This structured approach to maintaining the data dictionary ensures that all modifications are tracked and transparent, aiding in data management and project documentation.

V1.0

- Date of Change: 22/01/2024
- Description of Change: First dictionary made
- Reason for Change: Start of the project, iteration zero
- Impacted Fields: All items below have been added
    - Marketscreener
        - Date
        - Buy
        - Outperform
        - Hold
        - Underperform
        - Sell
    - Yahoo Finance
        - Date
        - Open
        - High
        - Low
        - Close
        - Adj Close
        - Volume