

Project Number: 31 项目编号: 31
Project Title: Ontology-Enhanced PCA Models Using Eurofidai ESG Dataset

项目名称: 利用 Eurofidai ESG 数据集建立本体增强型 PCA 模型使用 Eurofidai ESG 数据集的本体增强型 PCA 模型

Project Clients: Mingqin Yu

项目客户: 于明琴

Project specializations: Software Development;Computer Science and Algorithms;Big data Analytics and Visualization;

项目专长: 软件开发; 计算机科学与算法; 大数据分析与可视化;

Number of groups: 3 groups

组数: 3 组

Main contact: Mingqin Yu

主要联系人: 于明琴

Background: 背景介绍

The integration of ontology-driven methodologies with Principal Component Analysis (PCA) can significantly enhance ESG metric interpretation and risk assessment. The Eurofidai ESG dataset, containing 105 raw risk data metrics across global companies, provides an ideal dataset for this study. The primary objective of this project is to develop an ontologyenhanced PCA model that improves data structuring, dimensionality reduction, and interpretability in ESG analysis.

将本体论驱动的方法与主成分分析法（PCA）相结合，可以显著提高 ESG 指标的解释和风险评估能力。Eurofidai ESG 数据集包含全球公司的 105 个原始风险数据指标，为本研究提供了一个理想的数据集。本项目的主要目标是开发一种本体增强型 PCA 模型，以改进 ESG 分析中的数据结构、降维和可解释性。

This project aligns with the ontology-driven architecture proposed by Yu, Rabhi, & Bandara (2024), which emphasizes semantic standardization and structured ESG knowledge representation. Through this study, students will gain hands-on experience with machine learning, ontology engineering, and sustainability analytics.

本项目符合 Yu、Rabhi 和 Bandara（2024 年）提出的本体驱动架构，该架构强调语义标准化和结构化 ESG 知识表示。通过这项研究，学生将获得机器学习、本体工程和可持续发展分析方面的实践经验。

Requirements and Scope: 要求和范围:

2.1 Scope 2.1 范围

Phase 1: ESG Data Preprocessing and Ontology Integration

第 1 阶段: ESG 数据预处理和本体集成

This phase focuses on structuring the Eurofidai ESG dataset using ontology-based ESG categorization and preparing the data for PCA analysis.

这一阶段的重点是利用基于本体的 ESG 分类来构建 Eurofidai ESG 数据集, 并为 PCA 分析做好数据准备。

Task 1: Data Exploration

任务 1: 数据探索

Load and explore the Eurofidai ESG dataset, identifying key features, missing values, and anomalies.

加载并探索 Eurofidai ESG 数据集, 识别主要特征、缺失值和异常情况。

Conduct summary statistics and visual analysis to understand metric distributions and data variability.

进行汇总统计和可视化分析, 以了解指标分布和数据变化。

Task 2: Data Preprocessing and Normalization

任务 2: 数据预处理和标准化

Handle missing values using data imputation techniques.

使用数据估算技术处理缺失值。

Standardize numerical ESG metrics to ensure comparability across dimensions.

将 ESG 数值指标标准化, 以确保各方面的可比性。

Convert categorical ESG features into ontology-compatible representations.

将 ESG 的分类特征转换为与本体兼容的表征。

Task 3: Ontology Integration

任务 3: 本体集成

Utilize ESG ontology models from Yu et al. (2024) to categorize ESG metrics into Environmental, Social, and Governance (E, S, G) groups.

利用 Yu 等人（2024 年）的 ESG 本体模型，将 ESG 指标分为环境、社会 and 治理（E、S、G）组。

Use SPARQL queries to retrieve structured ESG data and enrich metric relationships.

使用 SPARQL 查询检索结构化 ESG 数据并丰富度量关系。

Map ESG indicators to ontology-based feature hierarchies for improved interpretability.

将 ESG 指标映射到基于本体的特征层次，以提高可解释性。

Phase 2: Principal Component Analysis (PCA) for ESG Data

第 2 阶段：ESG 数据的主成分分析 (PCA)

This phase applies PCA to extract the most relevant ESG features while integrating ontology-based knowledge to enhance explainability.

该阶段应用 PCA 提取最相关的 ESG 特征，同时整合基于本体的知识以提高可解释性。

Task 4: Applying PCA

任务 4：应用 PCA

Compute the covariance matrix and extract eigenvalues & eigenvectors.

计算协方差矩阵，提取特征值和特征向量。

Determine principal components (PCs) that retain the most variance in ESG data.

确定 ESG 数据中保留最大差异的主成分 (PC)。

Reduce the dimensionality of ESG metrics, ensuring minimal information loss.

降低 ESG 指标的维度，确保信息损失最小化。

Task 5: PCA Interpretation & Feature Engineering

任务 5：PCA 解释与特征工程

Visualize PCA results using scree plots, biplots, and heatmaps.

使用scree plots、biplots 和 heatmaps 对 PCA 结果进行可视化。

Interpret the significance of each principal component (PC) in ESG risk assessment.

解释每个主成分 (PC) 在 ESG 风险评估中的重要性。

Compare traditional PCA vs. ontology-enhanced PCA models for interpretability and feature selection.

比较传统 PCA 模型与本体增强 PCA 模型的可解释性和特征选择。

Task 6: ESG Risk Factor Analysis

任务 6：ESG 风险因素分析

Identify which ESG categories contribute most to principal components.

确定哪些 ESG 类别对主要成分的贡献最大。

Link PCA-extracted ESG factors to corporate sustainability risks and decisionmaking.

将 PCA 提取的环境、社会 and 治理因素与企业可持续发展风险和决策联系起来。

Validate findings using external ESG reports and industry benchmarks.

利用外部 ESG 报告和行业基准验证研究结果。

Phase 3: Model Implementation, Evaluation, and Reporting

第 3 阶段：模型实施、评估和报告

This phase implements, evaluates, and presents the ontology-enhanced PCA model through an interactive dashboard and technical report.

该阶段通过交互式仪表板和技术报告来实施、评估和展示本体增强型 PCA 模型。

Task 7: Model Validation and Comparison

任务 7：模型验证和比较

Compare ontology-enhanced PCA with traditional PCA to assess model accuracy and interpretability.

将本体增强型 PCA 与传统 PCA 进行比较，以评估模型的准确性和可解释性。

Evaluate performance using variance explained, clustering accuracy, and ESG metric influence.

利用解释的方差、聚类准确性和 ESG 指标的影响来评估性能。

Task 8: ESG Dashboard

任务 8：ESG 仪表盘

Develop a visual dashboard (Dash, Streamlit, or Tableau) to present key PCA-driven ESG insights.

开发一个可视化仪表盘（Dash、Streamlit 或 Tableau），以展示 PCA 驱动的关键 ESG 见解。

Task 9: Final Presentation and Documentation

任务 9：最终演示和文件

Prepare a presentation summarizing project outcomes and key findings.

准备演示文稿，总结项目成果和主要发现。

Submit documented code, datasets, and results in a reproducible format.

以可重复的格式提交有记录的代码、数据集和结果。

Required Knowledge and skills:

所需知识和技能

The system will be developed as a data processing and ESG analytics platform, integrating ontology-based knowledge representation with PCA-driven feature selection and dimensionality reduction. The following features and specifications define the software/system requirements:

该系统将作为数据处理和环境、社会 and 治理分析平台进行开发，将基于本体的知识表示与 PCA 驱动的特征选择和降维整合在一起。以下特征和规格确定了软件/系统要求：

A. ESG Data Processing and Ontology Integration

A. ESG 数据处理和本体集成

Feature Extraction 特征提取

Load and preprocess Eurofidai ESG dataset from multiple reporting sources.

从多个报告来源加载并预处理 Eurofidai ESG 数据集。

Handle missing data and ensure data consistency before PCA transformation.

处理缺失数据，确保 PCA 转换前数据的一致性。

Ontology-Based ESG Categorization

基于本体的 ESG 分类

Assign ESG metrics to predefined ontology-based categories (E, S, G).

将 ESG 指标分配给基于本体论的预定义类别（E、S、G）。

Retrieve structured ESG relationships using SPARQL queries from the ontology knowledge base.

使用 SPARQL 查询从本体知识库中检索结构化的 ESG 关系。

Link ESG metrics to industry-specific sustainability concepts.

将 ESG 指标与特定行业的可持续发展概念联系起来。

Data Transformation for PCA

PCA 数据转换

Convert categorical ESG attributes into ontology-enriched numerical values.

将分类 ESG 属性转换为本体丰富的数值。

Standardize ESG risk metrics for PCA processing.

将用于 PCA 处理的 ESG 风险指标标准化。

B. PCA-Based ESG Feature Selection and Risk Assessment

B. 基于 PCA 的 ESG 特征选择和风险评估

PCA Model Implementation

PCA 模型的实施

Compute eigenvalues, eigenvectors, and explained variance for principal component selection.

计算主成分选择的特征值、特征向量和解释方差。

Apply dimensionality reduction while maintaining ESG interpretability.

应用降维技术，同时保持 ESG 的可解释性。

Visualization & Interpretability

可视化和可解读性

Generate heatmaps, biplots, and component contribution charts for ESG insights.

生成热图、双曲线图和成分贡献图，以深入了解 ESG。

Link PCA results to ontology concepts to enhance explainability.

将 PCA 结果与本体概念联系起来，提高可解释性。

Implement anomaly detection using PCA to identify ESG reporting inconsistencies.

使用 PCA 实施异常检测，以识别 ESG 报告的不一致性。

C. ESG Dashboard C. ESG 仪表板

Interactive Dashboard 交互式仪表板

Develop a real-time visualization tool for PCA-based ESG analytics.

为基于 PCA 的 ESG 分析开发实时可视化工具。

Display principal components, risk scores, and ESG trend analysis.

显示主成分、风险评分和 ESG 趋势分析。

Expected outcomes/deliverables:

预期成果/交付成果：

Preprocessed ESG dataset with ontology-enhanced feature categorization.

经过本体增强特征分类预处理的 ESG 数据集。

PCA-based ESG factor analysis, identifying key sustainability risk drivers.

基于 PCA 的 ESG 因子分析，确定关键的可持续发展风险驱动因素。

Ontology-driven feature engineering applied to ESG data.

应用于 ESG 数据的本体驱动特征工程。

Python-based implementation, including Jupyter Notebook documentation.

基于 Python 的实现，包括 Jupyter Notebook 文档。

Comparative analysis of traditional PCA vs. ontology-enhanced PCA models.

传统 PCA 模型与本体增强 PCA 模型的比较分析。

Final project report & presentation, detailing methodology, findings, and business applications.

最终项目报告和演示，详细介绍方法、研究结果和业务应用。