PATENT APPLICATION

ML Feature Selection Algorithm for Enterprise Process Optimization

Patent Application No. 16/789,432

ABSTRACT

A system and method for automated feature selection in machine learning models applied to enterprise process optimization, comprising a multi-stage algorithmic approach for identifying, ranking, and selecting optimal feature sets from high-dimensional industrial sensor data and operational metrics. The invention employs a novel hybrid methodology combining statistical correlation analysis, information gain assessment, and recursive feature elimination within a distributed computing framework.

BACKGROUND OF INVENTION

[0001] Enterprise digital transformation initiatives frequently encounter challenges in identifying relevant features from large-scale operational datasets. Traditional feature selection methods often prove computationally intensive and yield suboptimal results when applied to heterogeneous enterprise data sources.

[0002] Existing approaches typically rely on singular methodologies such as correlation-based feature selection (CFS) or principal component analysis (PCA), which may fail to capture complex interdependencies in industrial process data.

SUMMARY OF INVENTION

[0003] The present invention provides a novel system and method for automated feature selection in machine learning applications, specifically optimized for enterprise operational data. The invention comprises:

(a) A distributed preprocessing framework for handling multi-modal sensor data

(b) A hybrid feature scoring algorithm incorporating multiple selection criteria

(c) An adaptive threshold mechanism for feature importance evaluation

(d) A recursive optimization process for feature subset selection

DETAILED DESCRIPTION

[0004] The invention operates through the following components and processes:

Data Ingestion Layer

[0005] The system accepts input from multiple data sources including:

- Industrial IoT sensor feeds
- Enterprise resource planning (ERP) system metrics
- Process control system logs
- Quality management system data

Feature Evaluation Engine

[0006] The core feature selection algorithm employs a three-stage evaluation process:

Primary Feature Scoring

- Statistical correlation analysis
- Information gain calculation
- Variance inflation factor assessment

Secondary Feature Validation

- Cross-validation performance impact
- Feature stability measurement
- Collinearity detection

Optimization Layer

- Recursive feature elimination
- Ensemble model validation
- Performance threshold verification

CLAIMS

A computer-implemented method for feature selection in machine learning models, comprising:

(a) Receiving operational data from multiple enterprise data sources;

(b) Performing distributed preprocessing of said data;

(c) Executing a hybrid feature scoring algorithm;

(d) Implementing recursive feature optimization;

(e) Outputting an optimal feature subset for model training.

The method of claim 1, wherein the hybrid feature scoring algorithm comprises:

(a) Statistical correlation analysis;

(b) Information gain assessment;

(c) Variance inflation factor calculation;

(d) Cross-validation performance evaluation.

A system for implementing the method of claim 1, comprising:

(a) A distributed computing infrastructure;

(b) A feature evaluation engine;

(c) An optimization module;

(d) A validation framework.

INVENTORS

Dr. Robert Martinez

Chief Innovation Officer

Summit Digital Solutions, Inc.

Michael Chang

Chief Technology Officer

Summit Digital Solutions, Inc.

ASSIGNMENT

The inventors hereby assign all right, title, and interest in this patent application to Summit Digital Solutions, Inc., a Delaware corporation having its principal place of business at 2200 Innovation Drive, Suite 400, Boston, MA 02110.

DECLARATION

I hereby declare that I am the original inventor of the subject matter which is claimed and for which a patent is sought; that I have reviewed and understand the contents of this application; and that all statements made herein of my own knowledge are true and correct.

Executed on: January 15, 2024

_

Dr. Robert Martinez

_

Michael Chang

ATTORNEY DOCKET INFORMATION

Attorney Docket No.: SDS-PAT-2024-001

Law Firm: Thompson & Bradley LLP

Attorney of Record: Sarah J. Thompson, Reg. No. 58,392

Address: 100 State Street, Boston, MA 02109