International Journal of Intelligent Computing and Information Sciences

https://ijicis.journals.ekb.eg/

A Proposed Approach for Production in ERP Systems Using Support Vector Machine algorithm

H. ElMadany *

Department of Computer science,
Faculty of Computer and
Information Sciences,Ain Shams
University,
Cairo, Egypt
hassanelmadany@cis.asu.edu.eg

M. Alfonse

Department of Computer science,
Faculty of Computer and Information
Sciences,Ain Shams University,
Cairo, Egypt
Marco_alfonse@cis.asu.edu.eg

M. Aref

Department of Computer science,
Faculty of Computer and Information
Sciences,Ain Shams University,
Cairo, Egypt
mostafa.aref@cis.asu.edu.eg

Abstract: *Using machine learning in Enterprise Resource and Planning systems, enable the organization to store, manage and analyze their data to obtain the right decisions and gain valued visions that were previously unimaginable. There are many industries and lines of business that contain massive volumes of data such as manufacturing, finance, healthcare…etc. This paper introduces a proposed approach for using machine learning in the production module in Enterprise Resource and Planning systems. This approach is considered a novel attempt to enable Enterprise Resource and Planning systems to perform an automatic or semi-automatic decision in critical issues in manufacturing. The proposed approach is adopted for recommending a novel combination of product raw materials using machine learning. It reduces the cost, time, and efforts to produce a new product design that will help the organization improve its profitability.*

Keywords: *Machine learning, big data analytics, ERP, Materials discovery, Material design, Production Forecasting*

1. Introduction

Machine Learning (ML) uses sophisticated algorithms to teach computers to learn from data. Aimed at extracting information and gaining insight from vast databases, it learns from previous computations to deliver accurate, reputable decisions and outcomes and has therefore carried out a major role in many fields. It manages a massive volume of data which could be structured, semi-structured and unstructured data. These massive volumes of data are considered Big Data. Enterprise Resource Planning (ERP) systems can be defined as a business management software that used to facilitate the business process cycle by integrating all business modules and functions like Human Resource (HR), Purchasing, Sales,

* Corresponding author: H. ElMadany
Department of Computer science, Faculty of Computer and Information Sciences, Ain Shams University ,Cairo, Egypt
E-mail address: hassanelmadany@cis.asu.edu.eg

Inventory Control (IC), Finance, Production, Customer Relationship Management (CRM), Accounting, …etc. into a single system so your business operates as a seamless one. It covers the business needs and challenges in an effective, efficient, and productive way [1]. ERP contains big data indicating the effective execution of business processes. Using ML in ERP system, enables organizations to store, manage and analyze their data to obtain the proper decisions and gain valued visions that were previously unimaginable. In the past, mainstream research in this area focused on the implementation of critical success factors, ERP upgrades, project management practices, etc. Future research is needed to explore the potentials of ERP systems to be linked to current technology like big data and ML to get accurate decisions. There are many industries and lines of business that contain massive volumes of data such as manufacturing, finance, healthcare…etc.

Referring to [1], effective and forward-looking data source selection according to exact goals is essential for the effective use of ML in development. Therefore, we can rely on CRISP-DM which stands for Cross Industry Standard Process for Data Mining [2]. Fig.1 illustrates the phases of CRISP-DM [2] which can be summarized as follows:

- Business understanding: Understand and define the problem and the project's objectives from the business perspective and derive the data mining objectives.
- Data understanding: Collect the initial data and get familiar with the present data to identify the data's quality and underlying structure to gain initial insights.
- Data preparation: Build the final dataset for the modeling step from the initial raw data by attribute selection, reduction, pre-processing, and data cleaning. This step might be performed several times before the requirements of the modeling step are met.
- Modeling: Select, compare, and demonstrate modeling techniques on the prepared data set and calibrate the needed parameters to optimal values.
- Evaluation: Evaluate the models with the data mining objectives as well as with the business objectives to decide whether the model can be adopted or must be refined. In this step, novel ideas, requirements, or objectives can occur.
- Deployment: Implement the designed model in manufacturing and integrate the knowledge created in existing processes and systems.
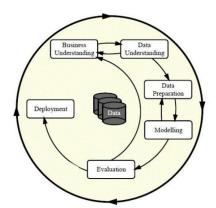

Figure 1: The Phases of the Current CRISP-DM Process Model for Data Mining [2].

Production in ERP provides manufacturing-specific functionalities that streamline and automate key business and production processes. The benefits of using ML in production could be briefly presented as follows [3]:

- Revolutionizing quality control through recognizing the root issues causing deficiencies and determining the possible risk factors for these deficiencies can significantly minimize waste and speed up production timelines.
- Predictively identifying issues, by reducing the existence of defects, may cause an enormous effect on the need for repair.
- Minimizing waste by making better determinations as to when equipment should be withdrawn from maintenance production.
- Increasing efficiency throughout the whole supply chain as ML help manufacturers to:
  - Predict how much and what kind of product they need.
  - Determine when they require it.
  - Know the most effective shipping route to transfer goods to their destination.
  - More accurately predict the potential problems that could slow down the supply chain.

This paper presents a proposed approach to suggest a combination of raw materials to produce a final product using ML techniques. This paper is organized as follows: Section 2 presents the background and related work. Section 3 presents the proposed approach and finally, the conclusion and future work are reported in section 4.

2. Background and Related Work

Repetitive experimental and theoretical characterization studies are often long-term and inefficient as a combination of raw materials and structures to generate a new product [4]. There is a relatively extended time to detect new material, usually about 10 to 20 years from initial research to first use. The authors in [5], divided the traditional methods to discover new materials into seven stages namely as follows: discovery, development, optimization, system design, certification, manufacturing, and deployment. It is properly recognized to be two traditional approaches widely used in the material science fields namely as follows: experimental measurements and computational simulation. Experimental measurements typically include microstructure and property analysis, calculation of properties, synthetic testing, etc. It often puts very high demands on facilities, the laboratory environment, and the researcher's expertise [4,6]. However, the computational simulation is the dynamic, process-focused model mounted on a computer that differs from conventional economic models to much more abstract concepts and processes [7]. Design materials using computational simulation will lead to less cost and time required by ML algorithms to identify patterns and the usage of algorithmic knowledge to model the linear or nonlinear relationships between material properties and related factors as ML designed to solve a materials science problem. The general paradigm of these learning systems can be represented as shown in equation (1).

$$Goal + sample + Algorith\,m = Model \tag{1}$$

Here, the goal is the question that has been defined and typically represented as an objective function. The sample is a population subset recommended to be analyzed. The Algorithm refers to an ML algorithm [8]. The selection of an appropriate ML algorithm affects the prediction accuracy and generalization ability. Each algorithm allows its scope of application, and thus, no algorithm is suitable for all problems. The commonly used ML algorithms in materials science can be divided into four categories namely as follows: probability estimation, regression, clustering, and classification.

Precisely, probability estimation algorithms are typically used for new materials discovery, whereas regression, clustering, and classification algorithms are used for material property prediction on the macro and micro-levels such as Support Vector Machine (SVM), Decision Tree (DT), Artificial Neural Network (ANN), and genetic algorithms. The Model can be defined as a description of a system in terms of mathematical concepts and refers to the algorithm that has been learned from the sample. The key concept behind ML in new materials exploration is to use a probabilistic model to test various structures and component combinations and finally pick the successful material from the candidate using Density Functional Theory (DFT) validation. ML typically focuses on empirical rules, which have been extracted by algorithms from a large amount of experimental knowledge [4,8]. Fig. 2 illustrates the general process of ML in materials discovery. It includes two core parts as below:

- Learning system: in which data collection, data preprocessing, model training, and testing are done.
- Prediction system: in which the system is prepared to predict new material.
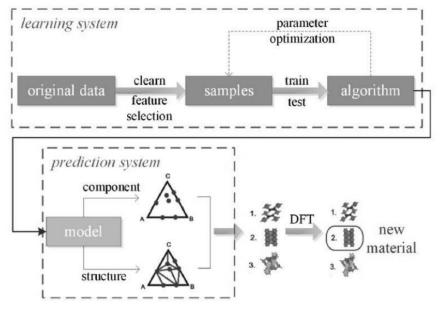

Figure 2: The General Process of Machine Learning in The Discovery of New Materials [4].

There are some research works in the field of chemical compositions that are likely to form compounds or discover new material components. The authors in [9] employed a Bayesian statistical system for extracting information from oxides experimental data available in ICSD which stands for the Inorganic Crystal Structure Database and predicting new ternary oxides successfully. Bayesian statistical models are implemented in the prediction of material components as they demonstrate good performance for the post-trust estimation. They have enormous importance due to the collection, mining, and processing of prior information. The authors in [10] used the Bayesian statistical system with the help of thermodynamic stability to accelerate the discovery process for novel ternary compounds instead of only ternary oxides. The computation time decreased by 6 magnitude orders in comparison with one calculation for first principles. The authors in [11] used ANN to discover unknown ionic liquids by designing new guanidinium salts. However, the authors in [12] used the failed reactions data entirely and proved an effective way of predicting reaction results for the crystallization of templated vanadium selenite using an SVM-derived DT algorithm trained on reaction results. The authors in [13] proposed a

model assessing the likelihood for ionic species to substitute for each other while preserving the crystal structure and proved how such a model can be used to predict new compounds and their crystal structures. Table 1 presents a comparison between the approaches discussed. The comparison was based on two criteria: the ML algorithm used, and the results achieved.

Table 1: Comparison between Machine Learning Techniques to Discover New Materials.

Reference	ML algorithm	Results
[9]	Bayesian	209 new compounds
[10]	Bayesian	4500 new stable materials
[11]	ANN	6 new guanidinium salts
[12]	SVM	89% success rate
[13]	Bayesian	substitution rates of 20 common ions

3. The Proposed Approach

The proposed approach of suggesting a combination of raw materials to produce a final product using ML techniques consists of four phases as below (Fig. 3 illustrates the proposed approach):

- Data collection phase.
- Preprocessing phase.
- Learning phase using the SVM algorithm.
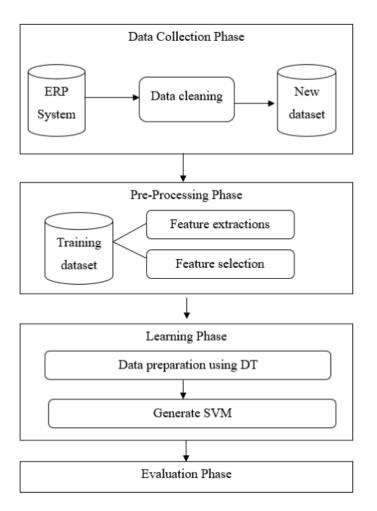- Evaluation (cross-validation) phase.

Figure 3:The Proposed Approach of Suggesting a Combination of Raw Materials to Produce a Final Product.

3.1 Data collection phase

The first phase is the creation of samples in materials science, obtaining the original data from computational simulations and experimental measurements [14]. This phase includes some processes such as:

- Selecting data: data collected from ERP system database. The collected data includes items, Bill Of Materials (BOM), routing stages, production process such as production orders and goods issue, sales data, purchasing data…etc.
- Data cleaning and preparation: data should be cleaned due to data incomplete, noisy, and inconsistent…etc.
- Constructing a new dataset: this process gets the output of the data cleaning process and constructs a new dataset containing the samples ready for training.

3.2 Pre-Processing phase

In this phase, the data collected from the previous phase is analyzed. This phase includes feature engineering which is a method for the extraction of features from raw data using data mining techniques through domain information. These features can be used to enhance the ML algorithm's performance.

3.3 Learning phase using the Support Vector Machine algorithm

This is the main step in our proposed approach in which learning is done. This phase links the input data to output data using a particular set of nonlinear or linear functions. Due to a unit-sensitive flaw in SVM [15] as different scaling methods will generate different results; the data will be adapted by the DT algorithm. DT does not require any knowledge or parameter setting. The data will be scaled based on the outcome that the decision tree will produce, so SVM can be adapted to the new data for training and prediction. This step will improve the generalization precision. Fig. 4 illustrates the pseudocode of the algorithm used in [15] to adapt the training data using the C4.5 algorithm [16] before using SVM.

Algorithm 1: Given the training samples

Step 1: Check for base cases.

Step 2: For each attribute a. Find the normalized information gain from splitting on a.

Step 3: Let a_best be the attribute with the highest normalized information gain.

Step 4: Create a decision node that splits on a_best.

Step 5: Recurse on the sublists obtained by splitting on a_best, and add those nodes as children of node.

Figure 4: The Pseudocode of the Algorithm used to Improve the SVM using DT [15].

The Decision Tree adapts a recursive top-down strategy. In an internal node, compare the attribute value, separate sub-sets by selected attribute into samples. As shown in figure (4), by giving the training set, the test set, and a user-specified constant μ, the C4.5 will be applied to the training set, generate the decision tree. For each dimension in both training set and test set, scale it to [0, 1] if it is not a node of outcome or scale it to $[0, \mu(L - l)]$ where l is the level of this node, and L is the total levels of the decision tree. Then we can use the SVM [17,18] to train the new training set and predict the test set.

3.4 Evaluation (Cross-Validation) phase

This is the last phase of our proposed approach. We will use the Hold-out method [19] which is the most basic form of cross-validation, the data collected is split into two sets: the train and the test dataset. The function approximator fits a function using the training set only. Then the function approximator predicts the output values for the data in the testing. The errors were collected to provide the average absolute test set error. The Classification Accuracy (CA) is used to evaluate models used in classification problems as shown in equation (2).

$$CA = S/N \qquad (1)$$

where S and N refer to the number of samples that are correctly classified and the total number of samples, respectively. Three main methods can be used to measure the result performances:

- **Mean Absolute Error (MAE):** measures the average magnitude of the errors in a set of predictions, without considering their direction. It is the average over the test sample of the absolute differences between prediction and actual observation where all individual differences have equal weight (see equation 3).
- **Root Mean Squared Error (RMSE):** is a quadratic scoring rule that also measures the average magnitude of the error. It is the square root of the average of squared differences between prediction and actual observation (see equation 4).
- **The Squared Correlation Coefficient (R^2):** It is a measure of the linear correlation (dependence) between two variables X and Y, giving a value between +1 and -1 inclusive (see equation 5).

$$MAE = \frac{1}{n}\sum_{j=1}^{n}\left[y_j - y_j^{\wedge}\right] \qquad (3)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(y_j - y_j^{\wedge}\right)^2} \qquad (4)$$

where y_j refers to *the observations,* $y_j^{\wedge}$ refers to the predicted values of a variable and

n is the number of observations available for the analysis

$$R^2 = \frac{\left(n\sum x_i y_i - \sum x_i \sum y_i\right)^2}{[n\sum x_i^2 - (\sum x_i)^2]\,[n\sum y_i^2 - (\sum y_i)^2]} \quad (5)$$

where y_i, x_i are the set of *observations* and n is the number of observations.

4. Conclusions and Future Works

In this paper, we proposed a novel approach for recommending a new combination of product raw materials using ML. The proposed approach will reduce the cost, time, and efforts to produce a new product design that will help the organization to improve its profitability.

Further research will concern with implementing the proposed approach, applying the proposed approach to real case studies such as the food industry, cosmetics, and so on, and finally evaluating and enhancing the proposed approach.

References

1. Stanula, Patrick, Amina Ziegenbein, and Joachim Metternich. Machine learning algorithms in production: A guideline for efficient data source selection. Procedia CIRP 78 (2018): 261-266.

2. Wirth R, Hipp J. CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining, 2000, p. 29-39

3. Machine learning in manufacturing, https://towardsdatascience.com/machine-learning-inmanufacturing-17c95290b1f6, last accessed 27/2/2021.

4. Liu, Yue, Tianlu Zhao, Wangwei Ju, and Siqi Shi. Materials discovery and design using machine learning. Journal of Materiomics 3(3) (2017): 159-177.

5. National Science and Technology Council (US). Materials genome initiative for global competitiveness. Executive Office of the President, National Science and Technology Council, 2011.

6. Wuest, Thorsten, Daniel Weimer, Christopher Irgens, and Klaus-Dieter Thoben. Machine learning in manufacturing: advantages, challenges, and applications. Production & Manufacturing Research 4(1) (2016): 23-45.

7. Prietula, Michael J., and Kathleen M. Carley. Computational organization theory: Autonomous agents and emergent behavior. Journal of Organizational Computing and Electronic Commerce 4(1) (1994): 41-83.

8. Kumar, S., and I. M. Nezhurina. Machine learning applications for design of new materials: A review. Industry 4.0 3(4) (2018): 186-189.

9. Hautier, Geoffroy, Christopher C. Fischer, Anubhav Jain, Tim Mueller, and Gerbrand Ceder. Finding nature's missing ternary oxide compounds using machine learning and density functional theory. Chemistry of Materials 22(12) (2010): 3762-3767.

10. Meredig, Bryce, Ankit Agrawal, Scott Kirklin, James E. Saal, J. W. Doak, Alan Thompson, Kunpeng Zhang, Alok Choudhary, and Christopher Wolverton. Combinatorial screening for new materials in unconstrained composition space with machine learning. Physical Review B 89(9) (2014): 094104.

11. Carrera, Gonçalo VSM, Luís C. Branco, Joao Aires-de-Sousa, and Carlos AM Afonso. Exploration of quantitative structure–property relationships (QSPR) for the design of new guanidinium ionic liquids. Tetrahedron 64(9) (2008): 2216-2224.

12. Adler, Phillip DF, Sorelle A. Friedler, Joshua Schrier, and Alexander J. Norquist. Machine-learning-assisted materials discovery using failed experiments. Nature 533(7601), (2016): 73-6.

13. Fujimura, Koji, Atsuto Seko, Yukinori Koyama, Akihide Kuwabara, Ippei Kishida, Kazuki Shitara, Craig AJ Fisher, Hiroki Moriwake, and Isao Tanaka. Accelerated materials design of lithium superionic conductors based on first- principles calculations and machine learning algorithms. Advanced Energy Materials 3(8) (2013): 980-985.

14. Peck RL, Olsen C, Devore JL. Introduction to statistics and data analysis. Cengage Learning, 2015.

15. Lin, Fuming, and Jun Guo. Improving support vector machine by preprocessing data with decision tree. In 2011 International Conference on Computer Science and Service System (CSSS), IEEE, 2011, pp. 467-469.

16. Quinlan, J. Ross. "Induction of decision trees." Machine learning 1(1) (1986): 81-106.

17. Lu WC, Ji XB, Li MJ, Liu L, Yue BH, Zhang LM. Using support vector machine for materials design. Advances in Manufacturing 1(2) (2013):151-9.

18. Wei J, Chu X, Sun XY, Xu K, Deng HX, Chen J, Wei Z, Lei M. Machine learning in materials science. InfoMat 1(3), 338-58 (2019).

19. Cross-Validation, https://www.cs.cmu.edu/~schneide/tut5/node42.html, last accessed 27/2/2021.