CHEMICAL ENGINEERING TRANSACTIONS
VOL. 103, 2023
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it
Guest Editors: Petar S. Varbanov, Panos Seferlis, Yee Van Fan, Athanasios Papadopoulos

ISBN 979-12-81206-02-1; **ISSN** 2283-9216
DOI: 10.3303/CET23103108

Generating Design Space for Production Planning Problems Using Machine Learning Algorithms.

Abhinay Sharthi, Sumit Kumar, Nitin Dutt Chaturvedi*

Process Systems Engineering Lab, Department of Chemical and Biochemical Engineering, Indian Institute of Technology Patna, Bihta,801106, Patna, Bihar India
nitind@iitp.ac.in

Sustainable production planning can be achieved by including several key factors along with production which affect the environment and economy. In the current competitive market scenario, it is highly essential to emphasize several objectives simultaneously. Planning for sustainable manufacturing must take crucial issues like carbon emissions, energy use, and production cost minimization into account. The methodology for creating a practical design space for manufacturing with capped carbon emissions, energy use, and production prices is proposed in this study. Data-driven machine learning approaches are implemented to classify feasible data points from the data set to create design space and a model which predicts feasible and not feasible data points. Algebraic equations are constructed using support vectors. Using the equation, whether the selected process path satisfies all of the established criteria can be determined. To demonstrate the practical applicability and goal of the suggested techniques, an example is provided. The models have achieved an accuracy of 90 % using SVM and 100 % accuracy using Random Forest.

1. Introduction

Production planning is very essential to satisfy demand and achieve sustainable development. In the current competitive market scenario, it is highly essential to emphasise several objectives simultaneously. The majority of works of literature are concerned with maximising profit, minimising production and distribution costs, minimising delivery times, and so on. Three crucial aspects that must be taken into consideration in order to achieve sustainable development are energy usage, production costs, and carbon emissions. The goal of several objectives may also be imprecisely defined, which makes it difficult for the planner to define the definite goals clearly. This occurs when the goal comprises linguistic variables (like approximately equal to, essentially greater than, etc.), and the deterministic values of goals cannot be determined.
In recent years, various research studies have explored the use of machine learning algorithms in the context of production planning. In their paper Multi-Objective Optimisation for sustainable production planning, Kumawat et al. (2021) proposed a fuzzy approach for production planning that considers multiple process options, which addresses decision-making uncertainty in situations where there is flexibility in the problem at hand. In the paper (Badhotiya et al., 2019), Santander et al. (2023) Approach to Improving Production Planning, a Bayesian framework is implemented on an industrially relevant fluid catalytic cracking process model and compared to the production planning process traditionally followed in the refining industry. The resulting production planning structure has robust performance due to considering uncertainty in model predictions. A study by Riazi et al. (2023) proposes a new feature selection method based on Support Vector Machine (SVM) technique that considers multiple environmental parameters such as waste, pollution, and energy. The aim is to predict sustainable productions with high demand and minimal environmental impact. Results show that the proposed hybrid method outperforms other similar methods in accuracy, enabling smart decisions for sustainable production. Mula et al. (2006) research emphasises the significance of taking uncertainty into account in production planning, as models that do so can result in less-than-optimal choices. It analyses the body of research on production planning in the face of uncertainty and offers a system of classification for these models. According to emission restrictions, a producer can choose between using conventional or green technology to make a single product, according to Hong et al. (2016). Analysis of the issue is done using cap-and-trade and

Paper Received: 15 April 2023; Revised: 03 June 2023; Accepted: 29 June 2023
Please cite this article as: Sharthi A., Kumar S., Chaturvedi N.D., 2023, Generating Design Space for Production Planning Problems Using Machine Learning Algorithms., Chemical Engineering Transactions, 103, 643-648 DOI:10.3303/CET23103108

obligatory emission-cap approaches. To improve the models, a dynamic programming approach is devised. Findings indicate that combining different technologies is only advantageous when emission caps are legally obligatory.
Overall the use of many machine learning techniques, such as the SVM kernel, has shown great promise in production planning. These techniques have the potential to significantly improve the efficiency of manufacturing industries by identifying the optimal production plan based on available data and minimising production costs. In this paper, the use of SVM, along with random forest classification, in the context of production planning. These experimental results demonstrate the effectiveness of these techniques for production planning problems and highlight their potential for improving manufacturing efficiency.

2. Problem Statement

Given a set of historical/simulated data with varying target loads of production plants operating using different process routes. Each targeted load will have its respective energy consumption, production cost, and emission information for carbon gas as a by-product. The energy consumption, production cost, and emission vary with different load conditions.
This paper aims to project the feasible region and create a model to predict the feasible routes of production satisfying the capped limits of energy consumption, production cost, and emissions of gases onto the space of production targets. Additionally, to generate an algebraic equation to attain objective-based optimisation in the feasible region.

3. Methodology

The methodology presented to solve the defined problem is elucidated in two sections. In Section 3.1 the basic mathematics linked with the SVM algorithm, which required a classify the data. In Section 3.2 the basic concept of working with a random forest classifier is explained.

3.1 Support Vector Machine (SVM)

SVM is a type of machine learning algorithm that is capable of performing various tasks such as classification, regression, and outlier detection in high-dimensional space. It achieves this by constructing a hyperplane or multiple hyperplanes that are farthest away from the nearest data points of each class. This leads to a clear separation between the classes, and the size of the margin between the hyperplanes is inversely proportional to the generalisation error of the classifier. In simpler terms, SVM creates a boundary between different classes of data by finding the most optimal hyperplane(s) with the maximum distance from the nearest data points of each class resulting in a high-accuracy classification.
Given a set of training vectors xi ∈ $\mathbb{R}^p$ i = 1,2….k in two classes, and an output vector O ∈ {-1,1}^{k,} the goal is to find w ∈ $\mathbb{R}$ and b ∈ $\mathbb{R}$ as a result of which the prediction made by sign($w^T\phi(x)+b$) is correct for the vast majority of samples. Support Vector Classification (SVC) solves the following primal problem:

$$min_{w,b,\varepsilon} \frac{1}{2} w^T w + C \sum_{i=1}^{k} \varepsilon_i; \text{ s.t.} \quad (1)$$

$$Oi\ (wT\phi(x)\ +\ b)\ \geq\ 1 - \varepsilon i \quad (2)$$

$$\varepsilon i\ \geq\ 0, \forall i \quad (3)$$

Intuitively, the goal is to maximise the margin (by minimising ||w ||^2=w^Tw). Also, when a sample is misclassified or falls within the margin border, a penalty is incurred. Ideally, the equation O_i ($w^T\phi(x)+b$) ≥1 would be true for all data, which indicates a perfect prediction because the issues aren't always fully separable with a hyperplane. Some samples are permitted to be separated from their proper margin boundary 'ε_i'. The penalty term C works as an inverse regularisation parameter, controlling the strength of the penalty. The dual of the primal problem

$$min_{\delta} \frac{1}{2} \delta^T Q \delta - e^T \delta; s.t. \quad (4)$$

$$O^T \delta = 0 \quad (5)$$

$$0 \leq \delta_i \leq C, \forall i \quad (6)$$

where 'e' is an all-ones vector and 'Q' is a (k×k) positive semidefinite matrix $Quv = zuzvK(xu,xv)$, where $K(x\text{v},xv) = \phi(xu)T\phi(xv)$ is the kernel. The terms δi are called the dual coefficients, and they are upper bounded by 'C'. This dual representation highlights that the training vectors are implicitly transferred into a higher dimensional

space by the function is highlighted by this dual representation ′ϕ′. Commonly used kernel functions are the polynomial kernel, radial basis function (RBF) kernel with bandwidth, and sigmoid kernel. The output of the decision function for a given sample 'x' becomes as follows and the predicted class corresponds to its sign.

$$\sum_{u \in SV} z_u \delta_i K(x_u, x) + b \tag{7}$$

It should be noted that the sum of the support vectors will consist of the samples that lie within the margin because the dual coefficients are zero for the other samples.

3.2 Random Forest Classifiers (RFC)

Random forest classifiers fall under the category of ensemble algorithms. Ensemble Technique is a technique that integrates several learning algorithms to produce the best predictive performance. Random Forest is a well-known algorithm to train early in the model development process, to see how the model performs and is very simple to build. In addition to simplicity, it provides a fairly sensible indicator of the importance it assigns to the options. Random forests function by constructing a multitude of decision trees at training time and giving the class that is either the mean prediction (regression) of the individual trees or mode of the categories (classification) as output. The random forest algorithm works in the following way:
Step 1: Starts by selecting arbitrary samples from the dataset.
Step 2: Then, decision trees are produced for the selected samples, and each decision tree provides us with a prediction result.
Step 3: Next, voting is carried out among the predicted results.
Step 4: At last, the outcome with the majority of votes is chosen as the final prediction.
The benefit of Random Forest is that it is among the most accurate classifiers. It is robust since several decision trees are involved in the process and overcome the problem of overfitting as the average of all the predictions is taken. It also handles values that are missing. Plus, the time required for computation is less since they use both structured and unstructured data. On the other hand, other techniques solely use structured data, so they need to keep their entire dataset as a training data set and use complicated methods for determination, leading to increased time usage. Some things that should be in mind so that Random Forest Classifiers give more accurate predictions are that features in the dataset should have enough predictive power. If the features aren't so strong, then the predictions would suffer. Secondly, the trees that form the forest and their resulting predictions should have low correlations with each other. This would enable all the trees to protect each other from their respective errors. If they are somehow highly correlated, then the trees may give errors in the same direction.

4. Illustrative example

In this section, the proposed methodology is demonstrated via an illustrative example. For example, the production planning for the Iron Steel industry is carried out via three available options, which are process routes in the industry. The current production scenario utilises existing process routes with fixed production limits. The example is solved using Python (version 3.11.3) on the computer having configuration: Intel(R) Core (TM) i5 (3 GHz) and 8 GB RAM.

4.1 Production planning of the iron and steel industry

In 2018, the IS sector of India was the second-largest in the world after China. India leads the production of sponge iron or Direct Reduced Iron (DRI) globally, along with being the third-largest finished steel consumer in the world after China and the USA. The coal-based route accounted for 79 % of total sponge iron production, which was 30.51 Mt in 2017–2018. There are mainly four classifications of steel production routes which are based on process routes in IS industry: (a) blast furnace/basic oxygen furnace (BF/BOF); (b) coal-based DRI-Electric Arc Furnace (Coal-based DRI-EAF); (c) gas-based DRI-EAF, and (d) scrap-EAF routes. In this case study, production planning through Blast Furnace/Basic Oxygen Furnace (BF/BOF), Coal-based DRI-Electric Arc Furnace (Coal-based DRI-EAF), and Scrap-EAF routes is considered. The assumption of the overall demand for crude steel is taken as 350 Mt.
A set of historical data could be useful and more accurate. However, for this study, simulated data is used to train the classification models. Simulated data comprises combinations of the loads that are to be produced using three different process routes to satisfy the overall demand for crude steel, and their respective carbon emissions, production cost, and energy consumption are calculated using equations from Table 1.
In Table 1, x is the t of the product produced using that process. The production cost, energy consumed, and carbon emission through a process route can be affected by many factors. So it need not be linear for this case

study; the above equations are assumed. A data set of 150 random production combinations is generated and divided into two categories: feasible (1) if the characteristics are below the capped value for the selected combination output target, otherwise not infeasible (-1). The criteria for classification are that the total production cost is 800 billion dollars, energy consumption is 19,000 GJ, and carbon emission is 8,000 t.

4.2 SVM classifier

The data presented above is utilised to train SVC to classify the data based on feasibility and forecast a viable region in the production space. Figure 2 depicts SVC's performance in estimating the viable region for crude steel manufacturing. A plane is produced in Figure 2(a) using a linear kernel that divides the feasible and non-feasible data points. The RBF kernel is used to build a plane in Figure 2(b). Table 2 summarizes the model's prediction result for an extra 100 points.

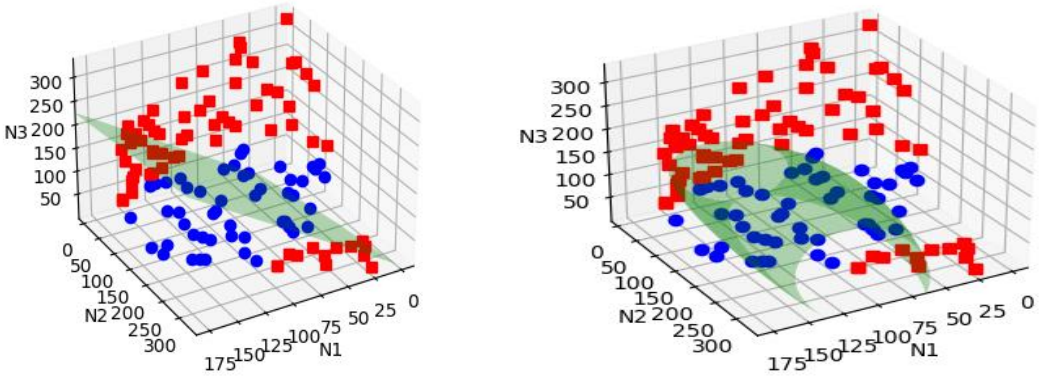

Figure 1: SVM using (a) Linear Kernel and (b) RBF Kernel

The mathematical equation of the plan generated by the linear kernel will be as ax + by + cz + d=0, which is a very simple equation that can be used to solve the problem.

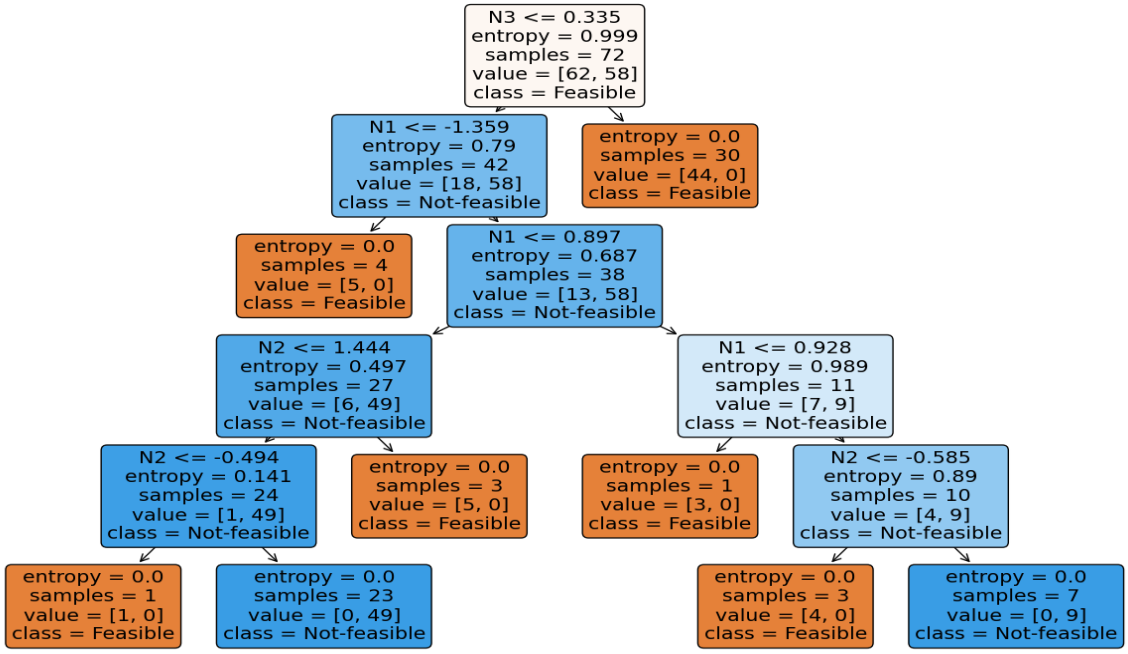

Figure 2: Decision tree of Random Forest Classifiers

4.3 Random Forest

To increase the accuracy of the prediction of feasible routes, a random forest classifier is implemented using 47 decision trees, which results in 100 % accuracy without overfitting. Figure 3 depicts the decision tree used for classifying whether a process route is feasible or not. Table 3 summarises the prediction result of an additional 100 points generated by the model. Precision can be seen as a measure of a classifier's exactness; recall is a measure of the classifier's completeness. The F1 score is a weighted harmonic mean of precision and recall such that the best score is 1.0 and the worst is 0. Support is the number of actual occurrences of the class in the specified dataset.

4.4 Tables

Table 1: Equations used for calculations

	Production cost ($)	Energy consumption	Carbon emission (CO_2)
Blast Furnace/Basic Oxygen Furnace (BF/BOF)(N1)	$29.275x^3-7115.8x^2+475202x+(5e+07)$	$0.0005x^3-0.1228x^2+8.2037x+855.48$	$(5e-05)x^3+0.0112x^2+0.7458x+77.771$
Coal-based DRI-Electric Arc Furnace (Coal-based DRI-EAF)(N2)	$97.206x^3-17076x^2+570426x+(1e+08)$	$0.0021x^3-0.3603x^2+12.036x+2133.5$	$0.0002x^3-0.0315x^2+1.0532x+ 186.68$
Scrap-EAF(N3)	$44.24x^3+8558.1x^2-380578x+(4e+07)$	$0.0016x^3+0.3032x^2-13.482x + 1250$	$0.0014x^3+0.2695x^2-11.984x+1111.1$

Table 2: Results of SVM

	Precision	Recall	F1-score	Support
Feasible (1)	0.90	0.90	0.90	20
Not Feasible (-1)	0.80	0.80	0.80	10
Accuracy			0.87	30
Macro average	0.85	0.85	0.85	30
Weighted average	0.87	0.87	0.87	30

Table 3: Results of Random Forest Classifiers

	Precision	Recall	F1-score	Support
Feasible (1)	0.87	1.00	0.93	20
Not Feasible (-1)	1.00	0.70	0.82	10
Accuracy			0.90	30
Macro average	0.93	0.85	0.88	30
Weighted average	0.91	0.90	0.89	30

5. Conclusion

This article offers a framework for production planning using a variety of process alternatives. The SVM model, which gives us a straightforward plane equation to solve, is used to address decision-making uncertainty in crisp problems. This is followed by the creation of a random forest classification model. An illustration of the aforementioned process is provided by the manufacture of crude steel. It was found that by altering a deterministic plan, it is possible to find a balance between carbon emission, energy consumption, and manufacturing cost. This paper demonstrates the approach's practical usefulness, and the suggested model outperforms the currently available remedy. Further research in this area can examine the viability of additional industry goals and incorporate more membership functions.

References

Badhotiya G.K., Soni G., Mittal M.L., 2019, Fuzzy multi-objective optimization for multi-site integrated production and distribution planning in two-echelon supply chain. The International Journal of Advanced Manufacturing Technology, 102, 635-645.

Chen X., Liao Z., Gao Z., Li, Q., Lv P., Zheng G., Yang K., 2022, A Calculation Model of Carbon Emissions Based on Multi-Scenario Simulation Analysis of Electricity Consumption. Sustainability, 14(14), 8765.
Dai X., Wang X., He R., Du W., Zhong W., Zhao L., Qian F., 2020, Data-driven robust optimization for crude oil blending under uncertainty. Computers & Chemical Engineering, 136, 106595.
Ehsani E., Kazemi N., Olugu E.U., Grosse E.H., Schwindl K., 2017, Applying fuzzy multi-objective linear programming to a project management decision with nonlinear fuzzy membership functions. Neural computing and applications, 28, 2193-2206.
Fang C.C., Lai M.H., Huang Y.S., 2017, Production planning of new and remanufacturing products in hybrid production systems. Computers & Industrial Engineering, 108, 88-99.
Gholamian N., Mahdavi I., Tavakkoli-Moghaddam R., Mahdavi-Amiri N., 2015, Comprehensive fuzzy multi-objective multi-product multi-site aggregate production planning decisions in a supply chain under uncertainty. Applied Soft Computing, 37, 585-607.
Hong Z., Chu C., Yu Y., 2016, Dual-mode production planning for manufacturing with emission constraints. European Journal of Operational Research, 251(1), 96-106.
Kumawat P.K., Sinha R.K., Chaturvedi N.D., 2021, Multi-objective optimization for sustainable production planning. Environmental Progress & Sustainable Energy, 40(6), e13741.
Mula J., Poler R., García-Sabater J.P., Lario F.C., 2006, Models for production planning under uncertainty: A review. International Journal of Production Economics, 103(1), 271-285.
Riazi A., Saraeian S., 2023, Sustainable production using a hybrid IPSO optimised SVM-based technique: Fashion industry. Sustainable Computing: Informatics and Systems, 37, 100838.
Santander O., Kuppuraj V., Harrison C.A., Baldea M., 2023, A Bayesian approach to improving production planning. Computers & Chemical Engineering, 173, 108226.