Self-Optimization of Reactive Sections for High Quality Production in Petroleum Refining Industry

By

Husnain Saghir

School of Chemical and Materials Engineering

National University of Sciences and Technology

2023

Self-Optimization of Reactive Sections for High Quality Production in Petroleum Refining Industry

Name: Husnain Saghir

Reg No: 00000328315

This work is submitted as an M.S. thesis in partial fulfillment of the requirement for the degree of

M.S. in Process Systems Engineering

Supervisor Name: Dr. Iftikhar Ahmad

School of Chemical and Materials Engineering (SCME)

National University of Sciences and Technology (NUST)

H-12 Islamabad, Pakistan

January, 2023

This thesis is devoted to my brother, my family, and my nephew.

Acknowledgment

All praise to "ALLAH," the architect of this world, who gave us the capacity for comprehension and sparked our curiosity about nature. Warmest welcomes to the "Prophet Mohammed (PBUH)," a source of knowledge and benefits for all of humanity.

Respect and sincere gratitude for my research **supervisor**, **Dr. Iftikhar Ahmad Salarzai** for his supervision and affectionate guidance. I would also like to extend my gratitude to **committee members**; **Dr. Tayyaba Noor** and **Dr. Nouman Ahmad**. for their valuable suggestions and guidance.

I am thankful to **Prof. Dr. Amir Azam Khan** (Principal School of Chemical and Materials Engineering) and **Dr. Erum Pervaiz** (HOD Department of Chemical Engineering) for providing a research-oriented platform.

In the end, I must express my very profound gratitude to my parents and my family. This accomplishment would not have been possible without them.

Husnain Saghir

Abstract

Hardware based sensing frameworks such as cooperative fuel research (CFR) engines are conventionally used to monitor research octane number (RON) in the petroleum refining industry. In this work, machine learning techniques are employed to predict the RON of two petroleum refining processes: (1) integrated naphtha reforming and isomerization process and (2) fluid catalytic cracking process. A dynamic Aspen HYSYS model was used to generate data by introducing artificial uncertainties in the range of ± 5\% in process conditions such as temperature, pressures, flow rates, etc. Generated data was used to train support vector machines, gaussian process regression, artificial neural networks, regression trees, and ensemble trees. Hyperparameter tuning was performed to enhance the prediction capabilities of gaussian process regression (GPR), artificial neural network (ANN), support vector machines (SVM), ensemble tree (ET) and regression tree (RT) models. Performance analysis of machine learning models indicates that in case of integrated naphtha reforming and isomerization process GPR, ANN, and SVM have R^2 values of 0.99, 0.978, and 0.979 and RMSE values of 0.108, 0.262, and 0.258, respectively. GPR, ANN, and SVM performed better than the remaining models and had the prediction capability to capture the RON dependence on predictor variables. ET and RT had an R^2 value of 0.94 and 0.89. In case of fluid catalytic cracking GPR, SVM, and ANN have R^2 values of 0.97, 0.969, and 0.963 and RMSE values of 0.3908, 0.3934, and 0.4333, respectively. ET and RT had an R^2 value of 0.941 and 0.88. The GPR model was used as a surrogate model for fitness function evaluations in two optimization frameworks based on the genetic algorithm and particle swarm method. Particle swarm optimization performed marginally better than genetic algorithm. The proposed methodology of surrogate-based optimization will provide a platform for plant-level implementation to realize the concept of industry 4.0 in the refinery.

Keywords: Machine Learning, Surrogate Modelling, Research Octane Number, Genetic Algorithm, Particle Swarm Optimization, Support Vector Machines, Gaussian Process Regression, Artificial Neural Networks, Regression Trees, and Ensemble Trees

Table of Contents

List of Figures

List of Tables

Nomenclature

Yi	Predicted value
Y	Expected value
i	Actual value
AI	Artificial Intelligence
ANN	Artificial Neural Network
ET	Ensemble Tree
GA	Genetic Algorithm
GPR	Gaussian Process Regression
KPIs	Key Performance Indicators
LHSV	Liquid Hour Space Velocity
MFFNN	Multilayer Feedforward Neural Network
ML	Machine Learning
RMSE	Root Mean Square Error
PSO	Particle Swarm Optimization
R^2	Coefficient of Determination
RON	Research Octane Number
RT	Regression Tree
SR	Straight Run
MAT	Micro Activity Test
SVM	Support Vector Machine
MSE	Mean Square Error

Chapter 1

Introduction

1.1 Background

In the recent era, there has been an ever-increasing growth of gasoline demand as a transportation fuel. The transport industry is a major contributor to greenhouse gases. Hence, research has been focused on the production of high-quality gasoline to achieve compliance with stringent ecological standards. According to the international council of clean transportation, 23% of global anthropogenic emissions of CO_2 are a result of the transportation industry. Road transport which utilizes gasoline as a primary fuel type is among the top contributors. CO_2 emissions can be restrained by designing high compression ratio engines which are in turn dependent on the octane rating of the fuels. Figure 1 shows the worldwide contribution of the transport sector to global CO_2 emissions [1].

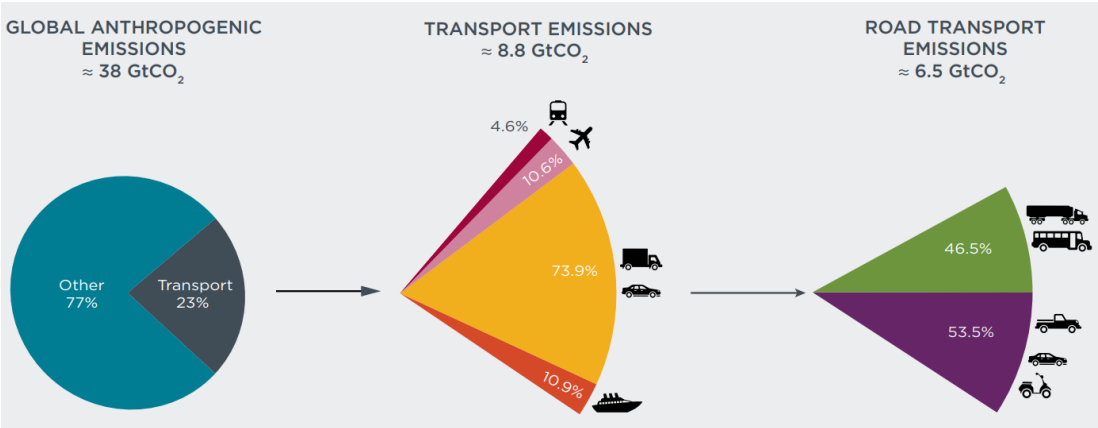

Figure 1 Contribution of the transport sector to global CO_2 emissions

Naphtha reforming, naphtha isomerization, and fluid catalytic cracking represent fundamental processes in petroleum refineries. These units are used to obtain gasoline with a higher value of octane number from raw naphtha. The isomerization process can maintain an allowable level of aromatics and benzene in gasoline.

The complete process flow diagram of the petroleum refinery is given in Figure 2 [2]. There are two types of reactive units in petroleum refineries: catalytic conversion processes and thermal conversion processes [3]. Chemical catalytic conversion processes include catalytic cracking, catalytic hydrocracking, catalytic reforming,

alkylation, isomerization, and hydro-treating. Thermal conversion processes include delayed coking and vis-breaking.

1.1.1 Catalytic Conversion Processes

In catalytic reforming low octane hydrocarbon in the naphtha is converted into high-octane gasoline without changing the boiling point of the component in the presence of a catalyst. The commonly used catalyst in the reforming section is platinum metal supported on silica or silica-based alumina. The main reaction is the dehydrogenation of naphthene to aromatics. The reaction is endothermic and requires a large amount of energy [4]. Hydrotreating unit catalytically stabilizes petroleum products by converting unsaturated hydrocarbons such as olefins and gum-forming diolefins into paraffin. Hydrotreating removes impurities such as sulfur, nitrogen, oxygen, and traces of halides from the feedstock. Hydrotreating is performed by reacting feedstock with hydrogen in presence of the catalyst. Different catalysts are used depending on the type of impurity needed to be removed. Cobalt and molybdenum oxide-based alumina catalyst is mostly used because they are easy to regenerate, highly selective, and have high resistance against poison [5]. Alkylation converts low molecular weight olefins to high molecular weight iso-paraffins by reacting with iso-paraffins. The reaction occurs at high temperature and pressure without a catalyst, typically 500 ^{o}C. An acid catalyst such as sulfuric acid and hydrochloric acid reduce the severity of operating conditions to a low-temperature range of 5 to 21^{o}C.

Isomerization enhances gasoline research octane number (RON) by converting straight-chain paraffin into a branch chain and simultaneously decreasing the benzene content through benzene saturation [6]. There are two isomerization processes, C_4 isomerization and C_5/C_6 isomerization. The C_4 isomerization process utilizes an aluminum chloride catalyst along with hydrogen chloride to produce feedstock for the alkylation process. A metal catalyst such as platinum is used in the C_5/C_6 isomerization process.

Breaking heavy oil into gasoline and a lighter product is called cracking. Cracking can be performed thermally or in the presence of a catalyst. The reaction is endothermic and takes place at a high temperature. The cracking result in the production of coke on the catalyst in case of catalytic cracking. Coke deposition minimizes catalytic activity and hence it is burned by using air at high temperatures. The process is classified based on fixed or fluidized beds [7]. Hydrocracking breaks heavy oil or vacuum residue into

valuable products such as gasoline, kerosene, diesel oil, and other light hydrocarbons by decreasing impurities and increasing the hydrogen-to-carbon ratio. The reaction occurs in a hydrogen-rich atmosphere at a temperature of 290-400 ^{o}C and pressure between 8275-13800 kPa [8].

1.1.2 Thermal Conversion Process:

Thermal conversion is used to convert residue of a high boiling point range into useful products or feedstocks for reactive sections in the refinery. Thermal cracking of vacuum distillation residue is achieved by using a delayed coking process. It is a cyclic process operating under the semi-batch operation. Coke, light gases, heavy gas oils,

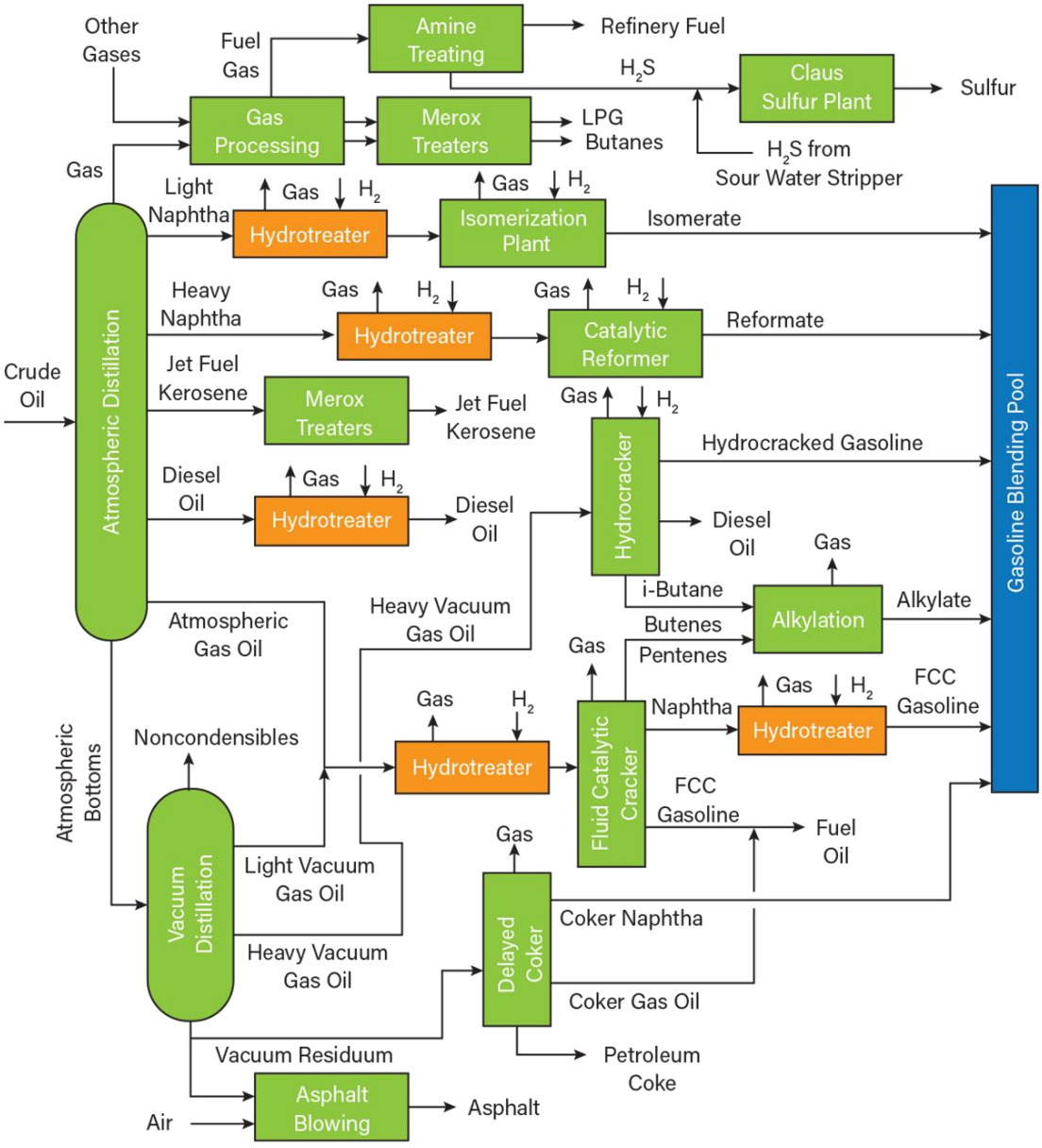

Figure 2 Petroleum refinery process flow diagram

and diesel are the end products of delayed coking. The mixture of feedstock and superheated steam is fed into the furnace and residence time is kept minimum to avoid coking in the furnace tubes. Cracking reaction takes place in the subsequent section of coke drums in which a combination of long residence times and sustained high temperatures results in cracking reactions. A coke bed is formed and cracked hydrocarbon products move to the fractionation column [9].

Visbreaking is utilized in the refinery to reduce the viscosity of vacuum distillate residue (VDR) and to increase the distillate yield in the petroleum refinery by producing light-end products and fuel oil. Visbreaking is a non-catalytic, mild thermal process that cracks large hydrocarbons by heating them in a furnace. Thermal severity and feed stock properties dictate the conversion of heavy ends to gasoline, light ends, and gas. Typically, 10-25 % conversion is achieved while producing the desired quality fuel oil. Visbreaking can be performed in two ways depending on the schematics of the reactors and where the cracking reaction takes place: coil vis breaking and soaker visbreaking. [10].

1.1.3 Quality Monitoring in the Refinery

Research octane number is an arbitrary scale defined based on two hydrocarbons namely iso-octane and n-heptane. RON of a fuel is defined as the volume percentage of iso-octane in a mixture of n-heptane and iso-octane that knocks with the same intensity as the fuel being analyzed. Under harsh conditions, iso-octane has high anti-knocking abilities and hence is assigned a 100-octane number. On the contrary, n-heptane shows the least resistance to autoignition and knocking and hence is assigned 0 RON [3].

Product quality monitoring and control in the continuous operation of the petroleum refining industry poses a challenging issue. Cooperative fuel research engine is traditionally used to measure the research octane number [11]. Near-infrared spectroscopy (NIRS) technology and gas chromatography are used to analyze the sample, measure the compositions, and estimate the research octane number [12, 13]. Hardware-based research octane number measuring systems require calibrations and have reliability and maintenance issues. Consequently, computational methods are being utilized in the recent era to predict the quality of refinery products.

Computational methods based on the black box or data-based modeling approach are being employed for quality improvement and monitoring. Data-based models are developed from input-output datasets and have the ability to predict the quality of gasoline while at the same time reducing computational costs. The data-based models are developed using statistical and soft computing-based techniques.

1.2 Objectives

1. Dynamic modeling of reactive units with the help of Aspen HYSYS and MATLAB by introducing artificial uncertainties of ± 5% in process conditions.
2. Development of ANN, GPR, SVM, regression tree, and ensemble trees for RON prediction.
3. Use of the machine learning method as a surrogate in the PSO and GA optimization framework for attaining high RON under uncertain process conditions.

1.3 Thesis Outline

This thesis is arranged as follows. Chapter 1 describes the background of the research work, followed by chapter 2 contains a detailed literature review. Chapter 3 discusses processes and applied machine learning models along with detailed research methodology to develop the hybrid framework. Chapter 4 contains the results and discussions about RON prediction and optimization along with the conclusions of the research work.

Chapter 2

Literature Review

Researchers have been focusing on the design and operation of reactive units of petroleum refining to realize sustainability. Computational methods have been employed to optimize the design and operation of the naphtha reforming and isomerization process and fluid catalytic cracking process. Studies can be broadly categorized into two groups, data based models and first principle models. First principle model is based on the knowledge of process.

Li et al. performed modeling of semi regenerative catalytic reformer by discretizing ordinary differential equations using the orthogonal collocation finite elements (OCFE) method. Results indicated that by integrating sparse matrix in successive quadratic programing (SQP) computational times can be reduced by 10 fold [14]. Stijepovic et al. developed a semiempirical mathematical model for catalytic reforming by using PONA analysis for kinetic lumping of equations. Simulation results were found to be in accordance with plant data [15]. Khosrozadeh et al. integrated mass and energy balances into a heterogeneous mathematical model for a semi-regenerative naphtha reforming process. Results showed an 0.3 % and 1.23% increase in RON and yield, respectively [16]. Kavousi et al. simulated the continuous catalytic reforming unit in PETROSIM software. REF-SIM module was utilized to perform reaction kinetics in PETROSIM. Sources of errors were identified for yield, and RON [17].

Chekantsev et al. developed a mathematical model to manage diverse feed compositions and reaction catalysts in the isomerization process and compared it with the experimental results [18]. Chuzlov et al. developed an Aspen HYSYS model for the isomerization reactor. An optimization mechanism was applied to increase the RON of the product by optimizing the composition of the feedstock [19]. The effect of feed composition and inlet temperature on yield and RON was investigated in subsequent research by implementing the mathematical models in IDE Delphi 7 [20]. Furthermore, efficiency estimation of isomerization processes was performed by utilizing chemical-technological system (CTS) modeling in Delphi 7 [21].

Said et al. utilized industrial data to perform calibration of isomerization reactors in Aspen HYSYS by identifying optimum feed conditions and temperatures, which resulted in heater duty reduction by 30% [22]. Buitrago et al. used the DWSIM software to simulate C5 isomerization and optimized the cooler and reactor temperatures [23]. Ahmed et al. worked on the kinetic parameters of naphtha isomerization reactors with industrial data and proposed modifications in the process [24]. Jarullah et al. compared the process developed by [24] to other competing industrial isomerization processes and found that the proposed isomerization configuration was superior in terms of cost, RON, and yield [25]. Mohamed et al. further validated the economic and product quality of isomerization process configurations. Results indicate that the work presented by [24] and [25] outperformed the existing models both in terms of cost and product quality [26].

First principle methods face problems like uncertainties, high dimensionality, nonlinearity, and time delays. In contrast, data based models offer an alternative strategy to achieve better accuracy and capability to deal with nonlinearities [27]. Various researchers have done data-based modeling of the naphtha reforming section.

Zahedi et al. trained two artificial neural network models by employing the radial basis function and back propagation methods to simulate a catalytic reforming unit. The volumetric flow rate of LPG, gasoline, and hydrogen, along with outlet temperature of reactors, Reid vapor pressure, specific gravity, and RON, was the target of the models. Results of the study indicated that the performance of the models was relatively high with a prediction error of 1.07% [28]. Sadighi et al. used recurrent layer artificial neural networks to predict the volume flow rate and RON of fixed bed catalytic naphtha reformer. An average absolute deviation of 0.238% and 0.813% for volume flow rate and RON of the gasoline, respectively, was observed [29]. Elfghi et al. performed a comparative study between artificial neural networks and response surface methodology (RSM) to simulate the catalytic naphtha reformer. Space velocity, pressure, and reaction temperature were used to predict the RON [30].

Sadighi et al. developed a hybrid decay function artificial neural network with recurrent layers to estimate the activity of Pt-Re/Al2O3 catalyst and predict the gasoline's RON and yield. The inlet temperature of reactors, hydrogen/hydrocarbon ratio, and liquid hour space velocity (LHSV) were the inputs of the model. A mean square error value of 0.0956 was observed in the prediction of gasoline RON and

catalyst activity was lost by 23% at the end of the cycle [31]. Al-Shathr et al. utilized a multilayer feedforward neural network (MFFNN) integrated with a mathematical model to estimate the functioning of the catalytic naphtha reforming. Performance indicators included temperatures, pressures, and weight fractions. ANN model performed better with the R^2 of 0.9467, 0.9403, and 0.9736 for pressure, mass fraction, and temperature, respectively [32]. Ahmad et al. developed a novel framework by integrating ANN and ensemble learning methods to estimate RON in the naphtha reforming process. Sensitivity analysis using the FAST and Sobol method were performed to investigate the individual effect of process parameter uncertainties [33].

Data based methods have also been used in the naphtha isomerization process. Herceg et al. used experimental data to develop data based models for the isomerization section by using SVM and parametric polynomial frameworks based on dynamic linear and nonlinear models [34]. Toch et al. estimated the kinetic parameters of the isomerization process using systematic kinetic modeling. A nonlinear regression model was developed using kinetic parameters to fit the process key performance indicators (KPIs). Results show an established correlation between the temperature and conversion of C_6 [35]. Sadighi et al. trained a hybrid artificial neural network to study catalyst deactivation, RON, and Reid vapor pressure (RVP) in an industrial-scale naphtha isomerization unit. Average absolute deviations of 0.0769% and 0.118% are observed for RON and RVP prediction [36]. Grozdani´c et al. fitted a linear model to experimental data to evaluate the isomerization process KPI's dependence on process parameters. One factor at a time (OFAT) and design of experiments (DOE) methods were used [37].

Various optimization strategies have been applied to naphtha reforming and naphtha isomerization processes. Wei et al. proposed a modified differential evolution algorithm for the optimization of the catalytic naphtha reforming process. Minimization of energy consumption and maximization of aromatics were the targets of optimization. The aromatic yield was increased by 2.34 % [38]. Zainullin et al. employed Non-dominated Sorting Genetic Algorithm II (NSGA-II) to perform multi-criterion optimization of catalytic reformer reactors. Benzene content was reduced by 4 % resulting in RON loss from 92.7 to 91.8 [39]. Babaqi et al. targeted reformate composition and RON by optimizing the kinetic and thermodynamic features by using particle swarm optimization in a MATLAB environment [40]. Pasandide et al.

performed global optimization of the catalytic reforming unit by utilizing the NOMAD algorithm to minimize energy costs. Aspen Plus and MATLAB were utilized to simulate the plant. An energy cost reduction of 13.5 % was observed [41].

Mencarelli et al. performed multistage optimization of the isomerization process by using a polynomial basis linear model. Linear model complexity was reduced in the first stage by using the subset selection method to identify the relevant basis function. In the second stage, an iterative optimization algorithm was used to optimize the LHSV, mass input, temperature, and pressure [42]. Duchene et al. performed kinetic model optimization by using the strategy developed by [42]. A quadratic model was used for the reforming reactor, and the numeric simulations of the reforming process were performed in Oscar 1.1. NLP derivative-free optimizer (SQL) was used for optimization purposes. The data based models reduced the computational time, but at the same time, the accuracy of prediction was also reduced [43]. Babaqi et al. coupled pinch analysis with particle swarm optimization to maximize yield and energy savings by modifying H_2/HC ratio, pressure, and temperature values. Energy savings of 16.2 % were achieved [44].

Investigation regarding the modeling and optimization of FCC units has been performed by various researchers. Sharma et al utilized the Aspen HYSYS process simulator and 21 kinetic lump method to simulate an industrial FCC unit using industrial operating parameters. The yield of gasoline, LPG, and LCO was the target variable of the study [45]. Yuming Zhang et al. used first principle modeling based on Aspen One simulation and kinetic lumped modeling to optimize the processing capacity and production yield of industrial FCC units. A yield increase of 0.5 % in gasoline and diesel was achieved [46]. Ahmed et al. performed dynamic modeling of FCC by simulating regeneration and riser section with the help of Simulink and Matlab models [47]. Yuhao Zhang et al. explored the ternary liquid equilibrium pathway to investigate the sulfolane, olefin, and sulfide interactions. Product composition was predicted using the UNIFAC contribution model. Regressive calculations were used to obtain the group interaction parameters along with Aspen Plus simulations to analyze the high olefin content of FCC components [48].

Tian et al. developed a hybrid framework of data driven and knowledge fusion-based approach for early warning and prediction of FCC abnormal conditions. The spearman ranking coefficient was used for dimension reductions of input data and to improve

deep learning model accuracy. The key variable prediction was performed using the convolution layer along with a long short-term memory network. The propagation path of abnormal conditions was deduced by utilizing a signed directed graph [49]. Chen et al. reduced the RON loss in the refining process by proposing random forest-based feature selection and a long short-term memory network for nonlinear mapping between yield and predictor variables. Furthermore, RON loss minimization was performed by using a gray wolf optimizer under the constraints of product sulfur content and RON loss reduction [50].

Santandera et al. employed deep learning frameworks by integrating production planning and economic model predictive control to reduce the economic gap between realized and planned production [51]. Yang et al. developed prediction frameworks for FCC product yields by integrating deep neural networks and kinetic lumped models. Hybrid models reduce the prediction error values by 9% by improving the correlations [52]. Long et al. proposed a hybrid structure for predictive model construction by combining LASSO and output-focused BPNN. LASSO was used for dimension reduction of input space. LASSO-BPNN algorithms were trained and tested on industrial data of FCC and comparison was performed with principal component analysis-BPNN and standalone BPNN models [53].

Ahmed et al. characterized the spent FCC catalyst by developing regression models and performing ANOVA on MS excel. surface area, bulk density, activity, and particle size were the regressed variables [54]. Abghari et al. trained a feed-forward ANN on industrial data for the prediction of gasoline quality, conversion, and LPG flowrates. Input variables of ANN included debutanizer column bottom temperature, feed flowrate, main fractionator top temperature, and reactor temperature. Furthermore, the firefly evolutionary algorithm was used to optimize the operating parameters. Chen et al. selected the relevant process data comprising property data of streams and operation parameters of FCC by employing an adaptive immune genetic algorithm (AIGA). Random forest was utilized for the development of FCC process models. A comparison of AIGA-RF was done with other hybrid modelling techniques [55]. Xinhe Chen et al. embedded surrogate models into the equation-oriented real-time optimization (RTO) framework of FCC. Trust region filter was employed for the selection of surrogate models and the Aspen RTO optimizer was replaced with a reduced model for FCC unit [56].

Various research works have been reported regarding naphtha reforming, isomerization, and fluid catalytic cracking units. However, a comparative analysis of machine learning methods applied to integrated naphtha reforming and isomerization units is not performed. Moreover, the application of machine learning methods as surrogate models in the evolutionary optimization environment is also not investigated.

Chapter 3

Process Description

3.1 Process Description:

3.1.1 Integrated Naphtha reforming and Isomerization Process

The process has two sections: naphtha reforming and the isomerization section. Integrated reforming and isomerization operations have the capability to transform naphtha feed mixture into high-quality gasoline with lower benzene content, hydrogen, and light hydrocarbon by-products. The naphtha section increases the RON, while the isomerization section decreases the benzene content in the end product. The naphtha reforming model contains 50 kinetic lumps and 115 reactions, and the isomerization model includes 20 kinetic lumps and 20 reactions. Figure 3 shows a schematic representation of an integrated naphtha reforming and isomerization process.

Naphtha Reforming. Naphtha feeds containing alkanes, a few aromatics, and cycloalkanes are mixed and fed to a splitter (SP-1) where light hydrocarbons below C4 are separated to stabilize the feed mixture. Stabilized feed is preheated in the heat exchanger (HX-1) to desired processing conditions and introduced into the distillation column (T-1) to separate light and heavy naphtha. Heavy naphtha is heated to the required temperature in the heat exchanger (HX-2) and sent to the reformer, while light naphtha is sent toward the isomerization section. Catalytic reformers operating at elevated temperatures (470-520^{o}C) and medium pressures (3-30atm) containing catalyst beds are arranged in series. Several reactions, including hydrocracking, hydrogenolysis, dehydrocyclization, and isomerization of alkanes, along with dehydrogenation and hydroisomerization of naphthenes and coke formation, occur in reformers. Temperature drops along the reactors as reaction is exothermic, so interstage heaters are introduced. Hydrogen is separated from the unstable reformer product and is sent to the isomerization section and storage. Unstable reformate is heated in HX-3 and is fed to the stabilization column (T-2) where light end gases are removed. Isomerate from the isomerization unit and stabilized reformate are blended into gasoline product.

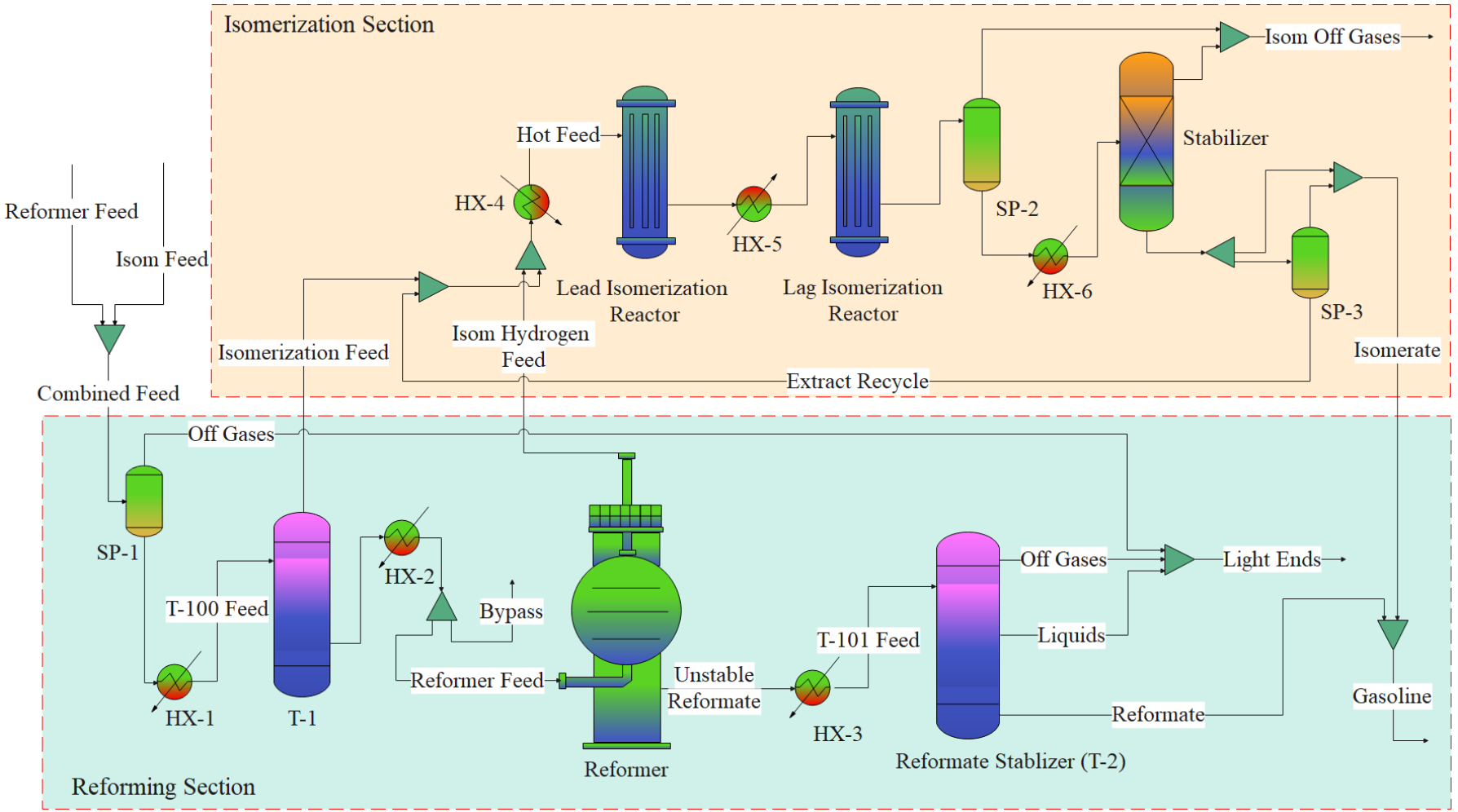

Figure 3: Process flow diagram of integrated naphtha reforming and isomerization process

Reactions: The reaction mechanism in process units is based on the components given in Table 1. Cycloalkanes and linear alkanes are converted to isoalkanes and aromatic having high RON. General reactions for components having i carbon atoms are as follows:

$$nP_i \rightleftarrows iP_i \quad (1)$$

$$nP_i \rightleftarrows N_i + H_2 \quad (2)$$

$$iP_i \rightleftarrows N_i + H_2 \quad (3)$$

$$N_i \rightleftarrows A_i + 3H_2 \quad (4)$$

Hydrogenolysis and hydrocracking of alkanes are according to the following secondary reactions:

$$P_i + H_2 \rightleftarrows P_j + P_k \quad (5)$$

P denotes a linear alkane or iso-alkane. The range of i is between 4 and 11 with i being equal to j + k.

Isomerization Section: The isomerization unit is a crucial process in the production of environmentally friendly gasoline. This process enhances the RON of C5/C6 naphtha by converting it to iso-alkanes while reducing the benzene content. Light naphtha from the top of the distillation column (T-1) is mixed with hydrogen from the reforming section and heated in the heat exchanger (HX-4). Feed enters lead and lag fixed bed isomerization reactors. Due to the exothermic nature of the isomerization reaction, le-chatlier principle dictates that increasing temperature decreases the yield of isoparaffins; hence reactors are operated at low-temperature values of 100-150^{o}C. Off gases, mostly hydrogen, are separated, and unstabilized isomerate is sent to the stabilization column, where light hydrocarbons are separated. Unconverted low octane number compounds are recycled back and mixed with feed.

Reactions: Reaction in the isomerization section proceed according to the following equation:

$$nP_6 \rightleftarrows 5P_6 \rightleftarrows 4P_6 \quad (6)$$

where P presents a linear alkane

Table 1: Compounds involved in Integrated naphtha reforming and isomerization process

Symbols	Compounds	Symbols	Compounds
H_2	Hydrogen	$8P_8$	n-octane
A_6	Benzene	A_7	Toluene
P_1	Methane	$6P_6$	n-hexane
P_2	Ethane	$8P_8$	n-octane
P_3	Propane	nP_9	Nonane
P_4	Butane	A_8	ethyl-benzene + xylenes
P_5	Pentane	iP_9	Isononane
A_7	Toluene	nP_{10}	Decane
$7P_7$	n-heptane	iP_{10}	Iso-decane
$5P_6$	simple branched alkanes with 6 C atoms	N_9	cycloalkanes with 9 C atoms
$4P_6$	double branched alkanes with 6 C atoms	A_9	9 C atom aromatics
N_6	cycloalkanes with 6 C atoms	N_{10}	cycloalkanes with 10 C atoms
$6P_7$	simple branched alkanes with 7 C atoms	A_{10}	aromatics with 10 C atoms
$5P_7$	double branched alkanes with 7 C atoms	nP_{11+}	hydrocarbons with more than 10 C atoms
N_7	cycloalkanes with 7 C atoms	iP_{11+}	isotopes with more than 10 C atoms
A_{11+}	aromatics with more than 10 C atoms	N_{11+}	cycloalkanes with more than 10 C atoms
$8P_8$	n-octane		

3.1.2 Fluid Catalytic Cracking

One of the key processes of petroleum refineries includes fluid catalytic cracking. Atmospheric distillation residues and gas oils from vacuum distillation are the feedstock of fluid catalytic cracking. Gasoline is the product of FCC with increased research octane number. The main sections of FCC units are the catalyst regenerator, riser, and fractionation tower. Feed is preheated to about a temperature of 490 ^{o}C and is atomized in the bottom section of the riser. The regenerated catalyst from the regeneration section is mixed with preheated feed to produce a temperature in the range of 530-570℃. Vaporized and preheated feed fluidizes the catalyst and pneumatically moves it to the top section of the riser. An endothermic cracking reaction takes place and the catalyst, and product are separated in disengaging vessels with the help of cyclones and deflector instruments. During cracking, coke deposits on the surface of the catalyst, and hence the activity of the catalyst is reduced. To regenerate the catalyst, the spent catalyst is moved to the regenerator where coke is burnt off by using hot air. Highly exothermic reactions of coke components such as hydrogen and carbon occur at about 690 ^{o}C in the regenerator section. The hot catalyst is then sent back to the bottom of the riser where it supplies heat for the cracking reactions. Reaction kinetics in the reactor is 21- lump system. Multiple reactions that are occurring in FCC are ring condensation, ring opening, cracking of paraffin, cracking of alkyl side chain, and coke formation. The schematics of the fluid catalytic process are shown in Figure 4 [57].

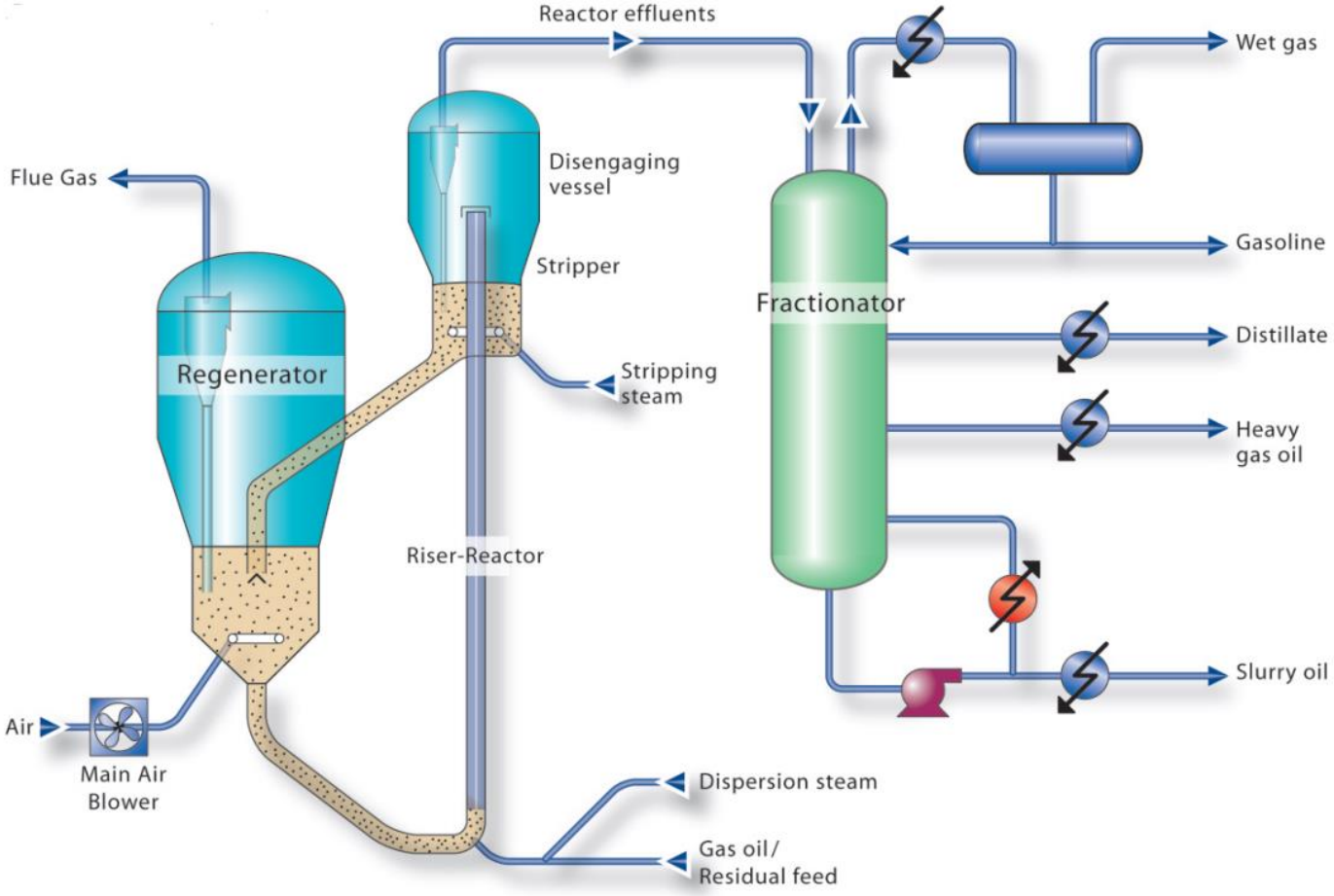

Figure 4 Process flow diagram of delayed cocking process

Catalyst: FCC catalyst is in the form of fine powder with 60-100 µm particle size, and bulk density in the range of 0.80-0.96 g/cm^3. Four major elements of FCC catalyst are zeolites, matrix, filler, and binder. Spherical particles of the micrometer size range are formed by mixing and spray-drying all the components. In addition to the above-mentioned components, additives can also be added to achieve specific purposes such as oil enhancing, NO_x and SO_x scavenging, poisonous metal trap, and promoter for CO combustion. A schematic presentation of the FCC catalyst is given in Figure 5.

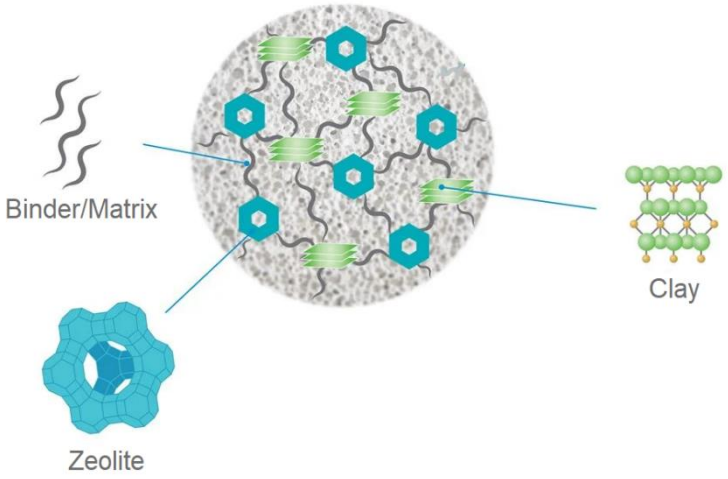

Figure 5 Fluid catalytic cracking catalyst

The key part of the FCC catalyst is the zeolite comprising 25% by weight of the catalyst. The zeolite used in the FCC catalyst is Faujasite. Zeolites act as strong solid acids by equating to 90 % sulfuric acid potency. The matrix component of the catalyst is made up of alumina and it is 40 % by weight of the catalyst. Matrix also contributes to the activity of the catalyst. FCC catalyst has a zeolite surface area (ZSA) and matrix surface area (MSA) of 150 and 200 m^2/gm. Integrity and physical strength are provided by the filler and binder part of the catalyst. Filler act as a transfer medium and heat sink by acting as the inner part of the matrix. The role of filler in catalyst activity is little. Kaolin is used as a filler material and silica solution is used as the binder.

3.2 Machine Learning Methods

Machine learning techniques applied in this thesis include gaussian process regression, support vector machines, artificial neural networks, regression trees, and ensemble trees.

3.2.1 Artificial neural networks

ANN constitutes a set of mathematical algorithms capable of estimating the relationship between huge amounts of data by using artificial neurons which imitate the biological functioning of the human brain. ANNs can act as surrogate models due to their ability to develop nonlinear correlations among data without the need for knowledge regarding the functioning of the system. In ANN structure, artificial neurons communicate with each other by using synaptic terminals arranged in layer form and further enhance the learning patterns of the algorithm by varying the weights. Accordingly, the algorithm consists of activity functions, weights, and biases. One of the commonly used ANN architectures for process modeling is a multilayer perceptron having three layers: input, output, and hidden. The number of variables in input and output data dictates the number of neurons in the input and output layer of the ANN structure. By increasing the number of hidden layers, the ability of ANN to estimate the complex relationship between data increases along with the chances of overfitting. A three-layered structure consisting of one hidden layer with sufficient neurons can determine the non-linear relationship between input and output pairs of data. ANN architecture is given in Figure 6.

$$Y_i = f(\sum_{j=1}^{N} X_j W_{ij}) \tag{7}$$

The activation function is given by:

$$\sigma(x) = \frac{1}{1+e^x} \tag{8}$$

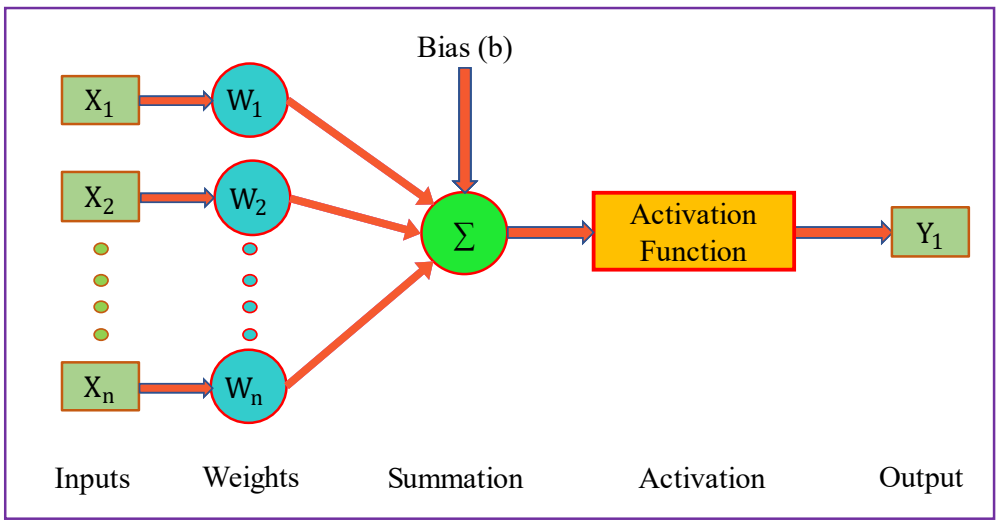

Figure 6 ANN architecture

3.2.2 Support Vector Machine

The SVM is a machine learning algorithm capable of function approximation that is employed for classification and regression. Kernel functions are utilized by SVM to develop input-output correlations. Kernel functions transform the data in such a way that non-linearity associated with the decision surface is linearized by transforming it into higher dimensions. Higher performance, simplicity associated with SVMs, and the ability to deal with small data sets make SVMs useful [24]. For training data set containing N samples of multivariate x_n with y_n as response value SVM generates linear regression function according to eq 3.

$$f(x) = x'\beta + b \tag{9}$$

b presents the bias and β is the weight value. Infeasible constraints are dealt with the introduction of slack variables for each point. The primal formula is constructed after the introduction of slack

variables [25].

$$J(\beta) = \frac{1}{2}\beta'\beta + C\sum_{n=1}^{N}(\xi_n + \xi_n^*) \tag{10}$$

Constraints:

$$y_n - (x_n' + b) \leq \varepsilon + \xi_n \tag{11}$$

$$(x_n' + b) - y_n \leq \varepsilon + \xi_n$$

$$\xi_n, \xi_n^* \geq 0$$

Box constraint is presented by *C*, a positive value that imposes the penalty margin on the observation that are outside the *ε* boundaries. ξ_n, ξ_n^* are slack variables indicating the error limits depending upon which side of ε tube sample lie. Lagrange dual formulation of SVM given in eq 6 results in computational simplicity [26].

$$f(x) = \sum_{n=1}^{N}(a_n - a_n')G(x_n, x) + b \tag{12}$$

a presents the Lagrange multiplier and $G(x_n, x)$ donates the kernel function

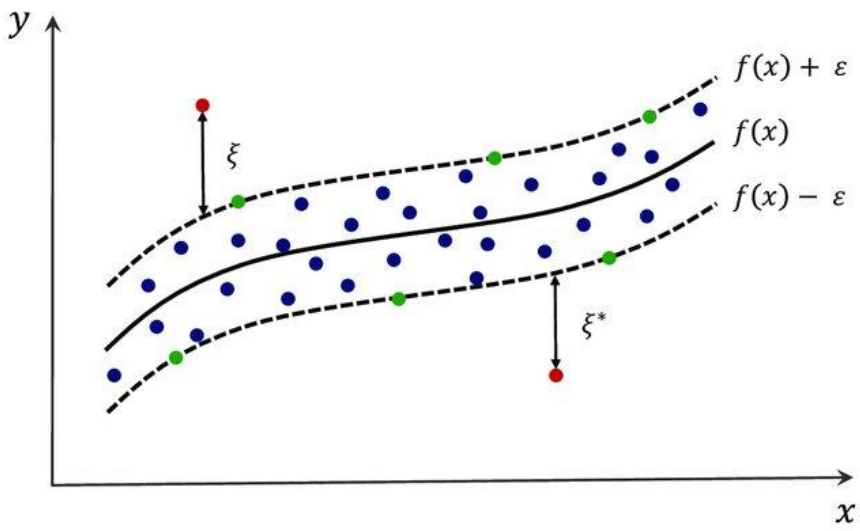

Figure 7 Support vector machine for regression

3.2.3 Gaussian process regression

GPR is a powerful nonparametric, nonlinear multi-dimensional regression tool used for classification and regression problems. GPR is based on Bayesian probability theory and is capable of interpolating and predicting data points scattered in input space. GPR instead of probability distribution calculation over a specific function performs calculations over all allowable functions to fit the data. Probability distribution functions are decided based on kernel functions [27]. Consider the training set $(x_i, y_i);\ i = 1,2, \ldots\ldots, n$, where $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$, taken from an unknown distribution. The GPR model gave the new input vector $x_{\text{new}}$and training data predicts the response value y_{new}. A linear regression model is of the form:

$$y = x^T\beta + \varepsilon \tag{13}$$

where $\varepsilon = N(0, \sigma)$ Data is utilized for the estimation of coefficients β and variance error σ^2. The response is elucidated by the GPR model by establishing latent variables, $f(x_i);\ i = 1,2, \ldots\ldots, n$, from a Gaussian process, and h as the basis function. The smoothness of response is captured by introducing the covariance function given in eq 8 and inputs x are projected into p-dimensional feature space by basis functions.

$$k(x_i, x_j) = \sigma_f^2\left(1 + \frac{\sqrt{5}r}{\sigma_l} + \frac{5r^2}{3\sigma_l^2}\right)\exp\left(-\frac{\sqrt{5}r}{\sigma_l}\right) \tag{14}$$

$$r = \sqrt{(x_i - x_j)^T(x_i - x_j)} \tag{15}$$

Where r is the Euclidean distance between x_i and x_j.

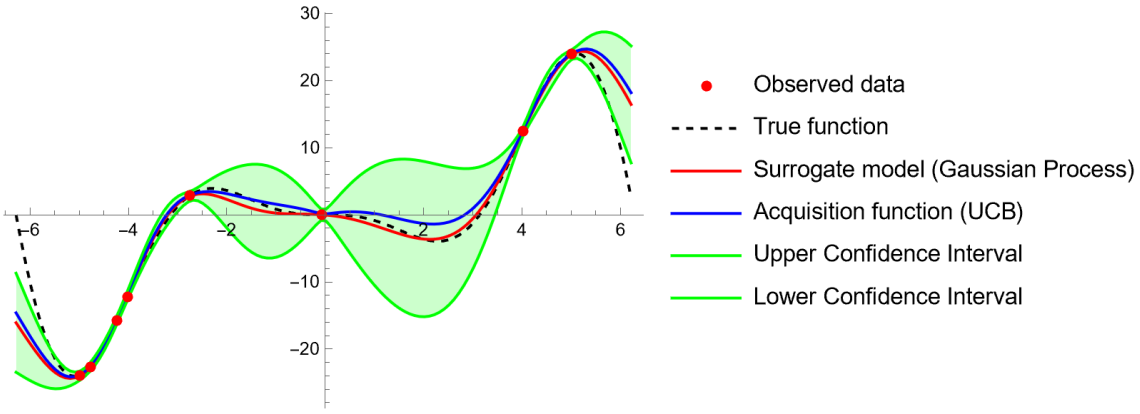

Figure 8 Gaussian process regression

3.2.4 Ensemble Trees

Ensemble tree is a predictive machine learning model that combines many weak learners to increase the prediction accuracy of the models. The ensemble of trees can be performed either by bagging or by boosting the regression trees. Averaging and bootstrapping methodologies are utilized in the bagging of regression trees to estimate the mean predicted values computed by all the individual trees. In boosting regression trees, data points are continuously updated, and wrong predictions are given extra weightage for the next models to correct them and reduce the bias. The model learns from the weak learners and develops a strong model [28].

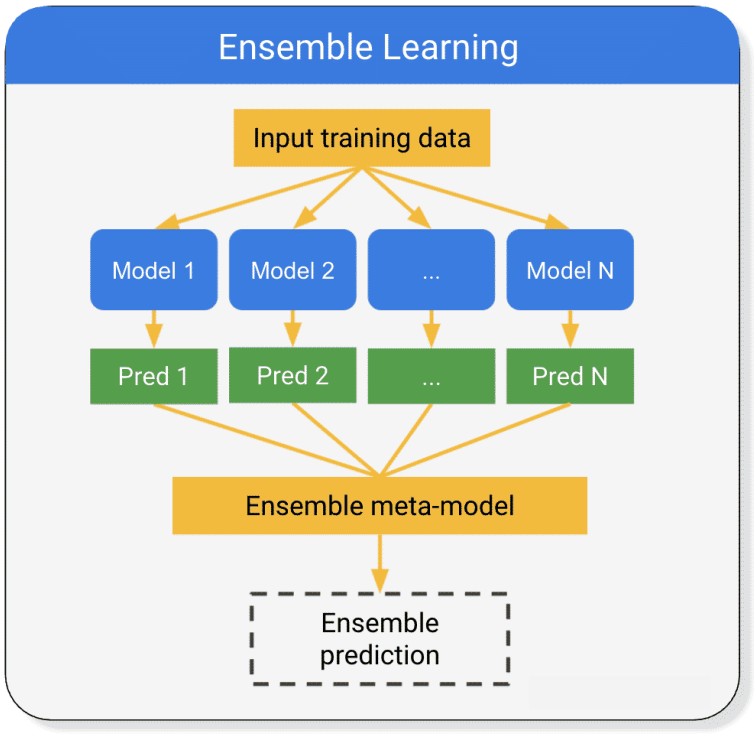

Figure 9 Ensemble learning architecture

3.3 Surrogate Based Evolutionary Optimization

Evolutionary algorithms solve optimization problems by evaluation candidate solutions based on their fitness function values. To obtain a satisfactory solution, many fitness function evaluations are required. In the case of real-world problems, the first principle model-based fitness function evaluations become computationally expensive and time taking. Surrogate based evolutionary optimization is utilized to avoid this inherent problem. In the case of surrogate based evolutionary optimization, fitness function evaluation is performed by utilizing robust approximations of first principle models by developing surrogate models. Surrogate models estimate the fitness function required for optimization.

3.3.1 Genetic Algorithm

The natural evolution of species based on the survival of the fittest is the guiding principle used in the formulation of genetic algorithms. The genetic algorithm investigates the response surface by using crossover and mutation operators in each iteration by utilizing a multi- point and parallel strategy. GA can optimize the process without the need for derivative information and initial guessing. GA uses a population-based strategy by indicating each parameter of the process by a gene and each solution by a chromosome. The objective function along with the decision variable range is defined. An iterative loop controlled by constraints and termination criterion is used to create a new population based on the favorable fitness function values. Offspring are created by the crossover operator by merging the chosen parents. Mutation operators change the fitness value of elements and if improved it is then passed on to the next generation. This process continues until the termination criterion is satisfied and an optimal solution is reached. The flowchart of the genetic algorithm is given in Figure 10.

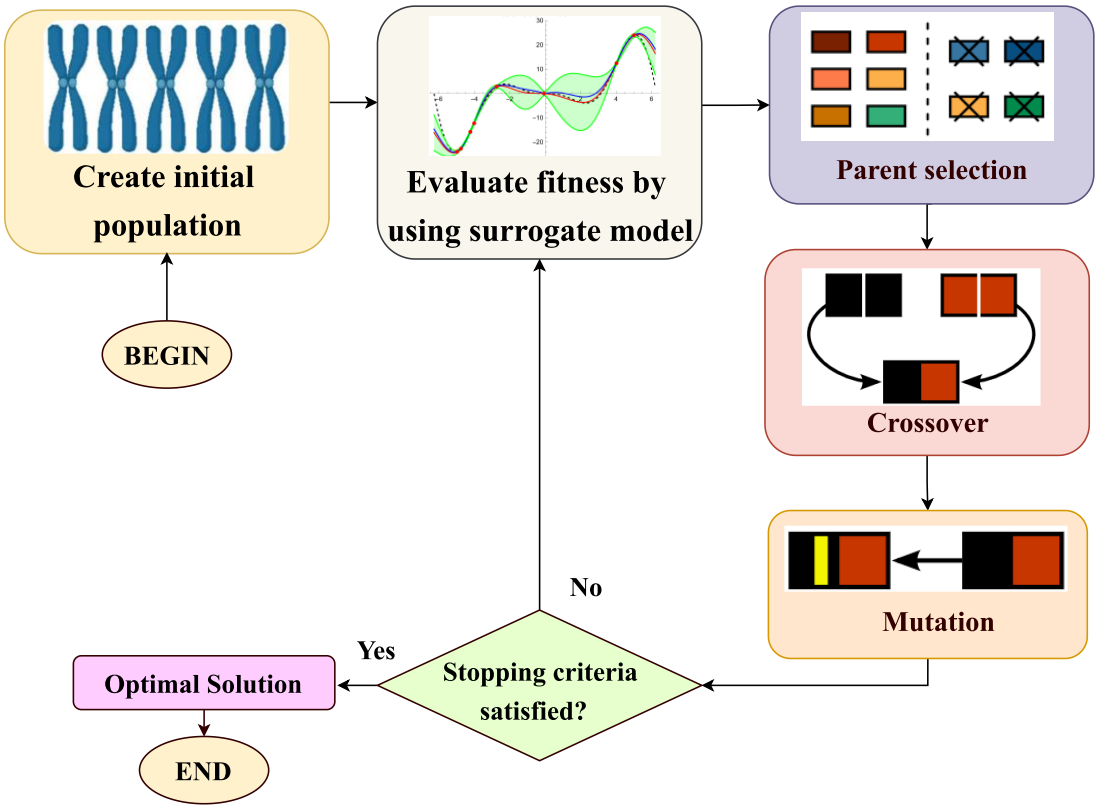

Figure 10 GA flowchart

3.3.2 Particle swarm optimization

Particle swarm optimization first developed by Eberhart and Kennedy is a stochastic, population-based evolutionary optimization technique [29]. The key idea behind the PSO algorithm is to reach the optimal position by guiding the swarm with the help of information sharing between the particles. In the search space, each particle or bird has its unique position and velocity. The best position of a particle (P_{best}) is calculated from the previous velocity of the particle and the best position discovered by the neighbors and the particle itself is known as G_{best}. P_{best} is based upon the best experience of a single particle while G_{best} is influenced by the neighbors in the swarm The algorithm keeps on calculating the new velocities and positions of the particles by changing the velocity location and neighbors until a specified stopping criterion is reached. Each particle has its velocity and position values in form of a vector in *n*-dimensional space. The movement of particles and swarm is dictated by the relationship between the range of the movement and the particle's position [30]. The velocity and position of any particle can be calculated by using equations 16 and 17.

$$V_{i+1} = WV_i + C_1 r_1 (P_{best} - X_i) + C_2 r_2 (G_{best} - X_i) \quad (16)$$

$$X_{i+1} = X_i + V_{i+1} \tag{17}$$

where P_{best} is the best position of ith particle, G_{best} is the best position among all particles. V_i is the velocity of ith particle, X_i is the position of the particle, W presents the inertial weight; C_1 and C_2 are acceleration coefficients, affecting the convergence of optimization. r_1 and r_2 are random numbers in the range of [0,1]. The flowchart of particle swarm optimization is given in Figure 11.

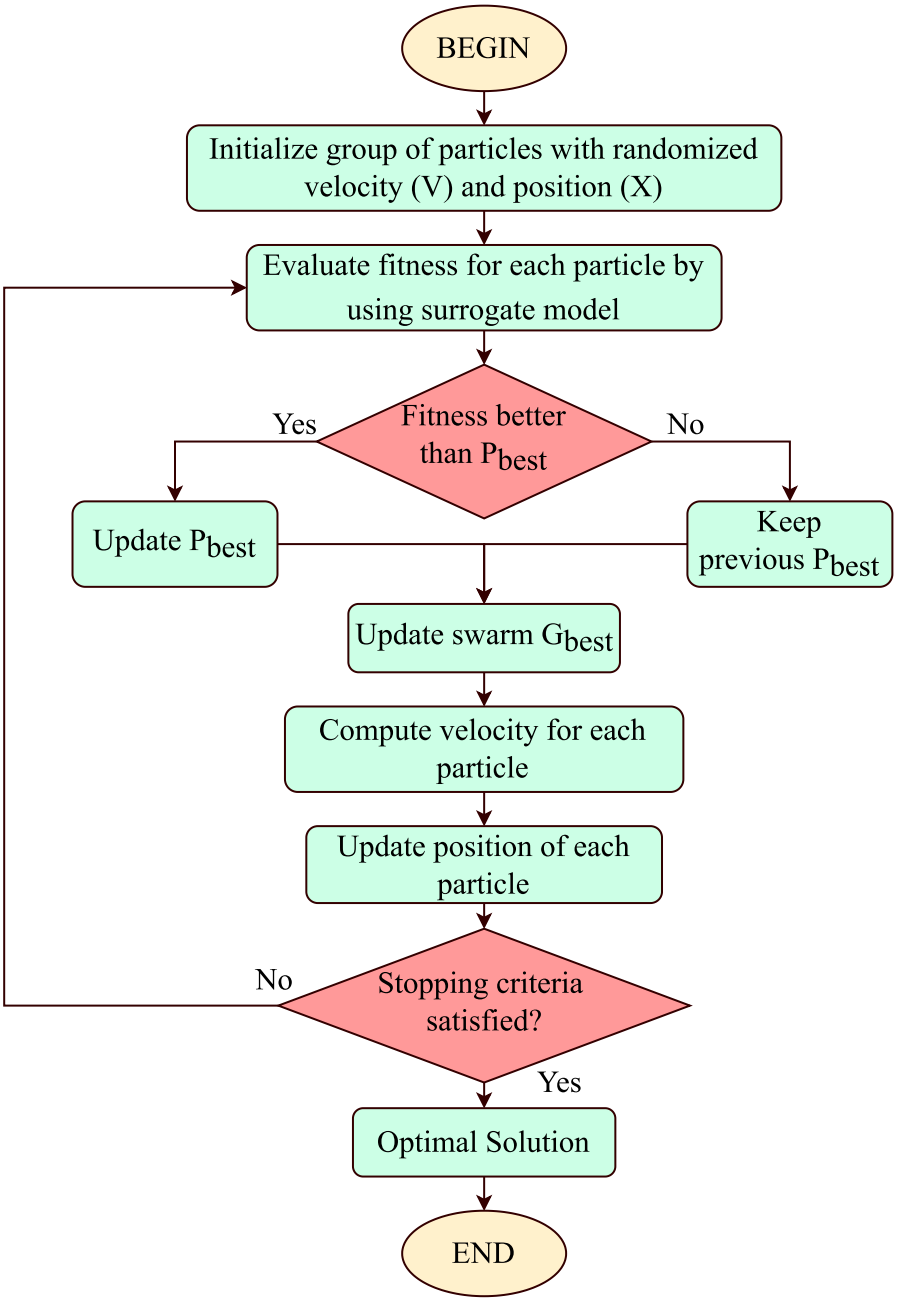

Figure 11 Particle swarm optimization flowchart

3.4 Methodology

Details of the multiple sections of the surrogate model-based evolutionary optimization approach are presented in Figure 12. Each of these sections is briefly discussed in the following paragraphs.

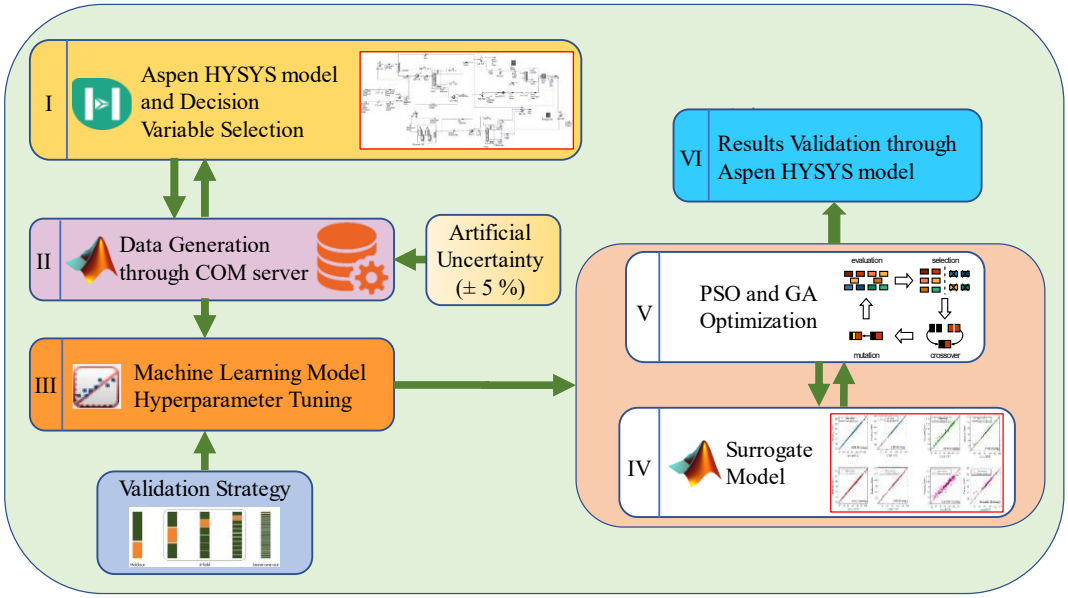

Figure 12 Methodology

Phase-I: First principle model and Decision variables

The first principle models of refining processes were developed from the Aspen HYSYS V10 and re-calibrated according to the data from the literature [58]. Predictor variables for the estimation and optimization of RON were selected by taking guidance from already published material on the individual sections of isomerization, reforming, and catalytic cracking [59].

Phase-II: Data Generation

The surrogate model requires accurate process data for training, validation, and testing purposes. In this study, MATLAB and Aspen HYSYS integration was used to generate data sets by introducing uncertainties in data in the range of ± 5%. To cover the response surface fully and effectively, data was generated randomly by creating an automated MATLAB script file interfaced with Aspen HYSYS using the COM server. Using this method, 360 input-output data sets were generated for the integrated naphtha reforming and isomerization process while 540 datasets were generated for the fluid catalytic cracking process.

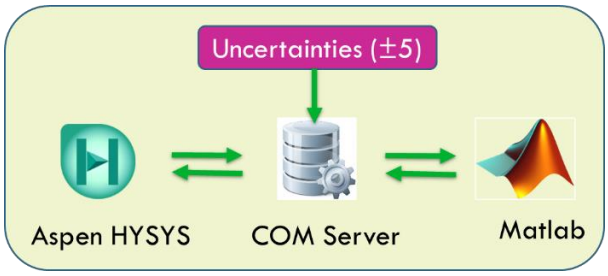

Figure 13 Data generation under uncertainty

Phase-III: Hyperparameter Tuning

Hyperparameter tuning is vital in increasing the prediction performance of machine learning models. A comparison of various combinations of hyperparameters was performed based on the robust K-fold cross-validation procedure. Overall datasets are split into training and testing sets of 80% and 20%, respectively. The data split was performed randomly by randomizing the indices of data sets. Hyperparameter optimization of each model was performed by splitting the training data sets into K folds. K-1 models were retained for training, while the remaining fold was used for validation. The process was done K times to ensure every fold was used once. Bayesian optimization was used for tuning hyperparameters along with the acquisition function expected improvement per second plus given by equation.

$$EI(x, Q) = EQ[\max(0, \mu Q(xbest) - f(x))] \quad (18)$$

100 iterations were performed to check the different combinations of hyperparameters. Mean Square Error (MSE) was used as a performance measuring criterion of different hyperparameter combinations. Hyperparameters resulting in the lowest MSE value are chosen for further model development.

Phase-IV: Machine Learning Model Development

The data sets from Aspen HYSYS's first principle model were normalized, and outliers were removed. 80% of the data sets were used for training and validation purposes, while the remaining were used for testing. Best hyperparameter combinations were used to develop GPR, SVM, ANN, ET, and RT using the training data sets. Then, models are applied to the test data to estimate the response variable value over unseen data. The prediction performance of machine learning models was evaluated based on R^2, MSE, and RMSE values calculated from equations:

$$MSE = \frac{1}{n}\sum_i^n (Y_i^{exp} - Y_i)^2 \quad (19)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_i^n (Y_i^{exp} - Y_i)^2} \quad (20)$$

$$R^2 = 1 - \frac{\sum_{i=0}^n (Y_i^{exp} - Y_i)^2}{\sum_i^n (Y_i^{exp} - Y_{avg}^{exp})^2} \quad (21)$$

Phase-V: PSO and GA Optimization

PSO and GA-based evolutionary optimization of the surrogate machine learning models was carried out using MATLAB's global optimization toolbox. The critical step in this strategy is that the fitness function in the case of genetic algorithm and particle swarm optimization was dictated by the trained surrogate machine learning models. The algorithm adopted for GA is as follows:

1. Constant parameters of GA were determined, such as the population of initial generation, the number of decision variables, the probability of crossover, and mutation.
2. Initial population creation using normalized responses, each response, or the "chromosome", contains eleven normalized values or "genes" corresponding to each of the eleven decision variables
3. Trained surrogate models were used to determine the fitness values of responses. The fitness value is then used to evaluate and rank different responses, and parents for the next generation are selected.
4. Mutation and crossover operators of the GA algorithm were used to create new responses.
5. Newly generated responses were evaluated based on their ranked fitness function value. Best responses formed the next generation.
6. Steps 4-6 are repeated until the termination criterion is satisfied. The highest-ranking response after the optimization termination was RON's best value.

The algorithm steps for PSO are as follows:

1. The initial population of particles was generated containing velocity and position vectors.
2. The surrogate model was used to evaluate the particle's position.

3. If a particle’s present position is better than its previous best position, update it.
4. Find the best particle (according to the particle’s last best places).
5. Update the velocities and positions of the particles.
6. Repeat steps 2-5 until the stopping criterion is satisfied.

Phase-VI: Results Validation

Values predicted and optimized by GPR, SVM, ANN, ET, and RT were then fed into the Aspen HYSYS first principle model to validate the results.

Chapter 4

Results and Discussion

4.1 Quality Prediction of Integrated Naphtha Reforming and Isomerization Section

Table 7 contains the steady state values and sample datasets containing the artificial uncertainties of ± 5\% introduced in the input parameters.

Section 4.1.1 summarizes the results regarding the hyperparameter optimization of GPR, SVM, ANN, RT, and ET. Section 4.1.2 focuses on the comparative analysis of ML models used to predict the RON of integrated naphtha reforming and isomerization processes. In section 4.1.3, SHAP analysis, along with partial dependence plots, are discussed. Surrogate modeling coupled with optimization strategies is given, and optimization techniques are compared in section 4.1.4.

4.1.1 Hyperparameter Optimization of Prediction Methods

The performance of ML methods is strongly dependent on the internal or hyperparameters of the models. Several hyperparameters of the models are selected and optimized to cater to the demands of specific prediction problems. Instead of manually selecting the hyperparameters, the regression toolbox of MATLAB v2022a was used. Bayesian optimization was employed to optimize and select specific hyperparameters of each model. 10-fold cross-validation was used to train and validate the performance of models for a specific set of hyperparameters. MSE was used as the selection criteria for hyperparameters, and optimization was performed, then, the hyperparameters giving the lowest MSE value were selected for further model development. Table 2 contains the hyperparameter search space along with optimized values for all models. Hyperparameter values of GPR were Isotropic Matern 3/2 kernel function with kernel scale of 188.1225 and sigma value of 4.1997. Hyperparameters of ANN included 3 hidden layers, a sigmoid activation function, and a lambda value of 3.76 e-08. Linear kernel function with box constraint value of 5.066 and epsilon value of 0.0243 was the hyperparameters of SVM. Bagging of regression trees, minimum leaf size of 2 and 12 number of learners were the optimized hyperparameters of the ensemble tree model.

Table 2 Tuned hyperparameter values of ML methods

Methods	Hyperparameters	Optimized Values	Hyperparameter Search Space
GPR	Basis function Kernel function Sigma	Linear Matern 5/2 0.052159	Constant, Zero, Linear Non/iso-tropic Matern 3/2. Rational Quadratic iso-tropic Exponential, , Non/iso-tropic Matern 5/2 0.0001-20.1131
SVM	Kernel function Box constraint Epsilon	Linear 433.7886 0.098674	Gaussian, Linear, Quadratic, Cubic 0.001-1000 0.0015295-152.9515
ET	Ensemble method Minimum leaf size Number of learners Number of predictors to sample	Bag 1 500 9	Bag, LSBoost 1-146 10-500 1-11
RT	Minimum leaf size	3	1-146
ANN	Layers Neurons in layers Activation Function In hidden layer	1 10 Sigmoid	1-3 1-100 Sigmoid, Tanh, None

4.1.2 Prediction Performance Evaluation

The research octane number of combined reforming and isomerization process was predicted by training machine learning algorithms such as ANN, GPR, SVM, ET, and RT. Optimized hyperparameters were used to train the models. MSE, RMSE, and R^2 parameters were employed to compare the estimation capabilities of the above-mentioned methods.

Values of training and testing RMSE, MSE, and R^2 are given in Table 3. RMSE has a non-negative value, and a lower value is an indication of better model prediction

performance. The coefficient of determination value ranges from 0 to 1, 0 being the value at which the output variable cannot be predicted from the regressor variable, and 1 means that the response variable is fully predictable from regressor variables. Results point out that GPR, SVM, and ANN with R^2 values of 0.99637, 0.97925, 0.97865, MSE values of 0.01168, 0.06679, 0.06871, and RMSE values of 0.10807, 0.25845, 0.26212 respectively indicate high performance for prediction of RON.

Table 3 Prediction performance of ML methods

	R^2		**MSE**		**RMSE**	
Model	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
GPR	0.99856	0.99637	0.00555	0.01168	0.07452	0.10807
ANN	0.99675	0.97865	0.01251	0.06871	0.11186	0.26212
SVM	0.98852	0.97925	0.04422	0.06679	0.21029	0.25845
ET	0.93399	0.94160	0.25424	0.18797	0.50422	0.43355
RT	0.86469	0.89068	0.52113	0.35183	0.72189	0.59315

The performance of regression and ensemble trees is on the lower end of the spectrum with R^2 values of 0.94160 and 0.89068, respectively. Regression and ensemble trees performed poorly and could not capture the nonlinearity associated with the RON prediction. RMSE of the regression tree and ensemble of trees was 0.59315 and 0.43355, respectively. The performance of the individual regression trees is lower as compared to the ensemble of trees. Bagging the trees increases the performance by averaging out the individual performance of the trees. Figure 14 shows the plots between actual and predicted values of RON for ANN, GPR, SVM, regression trees, and ensemble trees in terms of training and testing data. Scatter plots further solidify that GPR, SVM, and ANN are the better surrogate model for capturing the nonlinearity associated with the dependence of RON on the predictor variables.

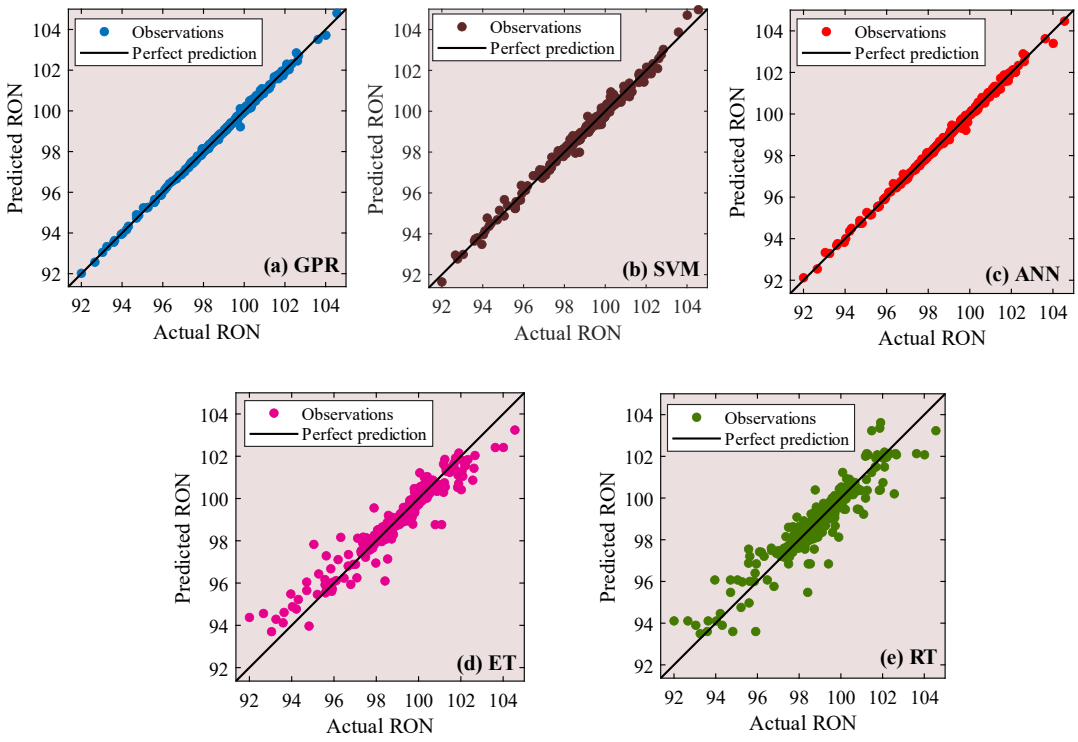

Figure 14 Actual vs predicted value of RON for (a) GPR, (b) SVM, (c) ANN, (d) ET and (e) RT

4.1.3 Partial Dependence Plots

SHapley Additive exPlanations (SHAP) is used to comprehend the effect of different predictor variables on the output performance of machine learning models. By analyzing the prediction performance and RMSE values of all the trained machine learning models it was concluded in the last section that GPR outperformed the other methods in handling the intricate relationship between RON and predictor variables. The impact of predictor variables on the RON value of the GPR model was interpreted by using shapley analysis. The shapley explanation chart given in Figure 15 shows the top 6 contributing variables to the output prediction of RON. The weighted average inlet temperature of the reformer and reformer feed flow rate are the positive contributing factors, while the isomerization feed flow rate was a negative contributing factor. The sequence and amount of impact of predictor variables on RON value are validated from the literature [33].

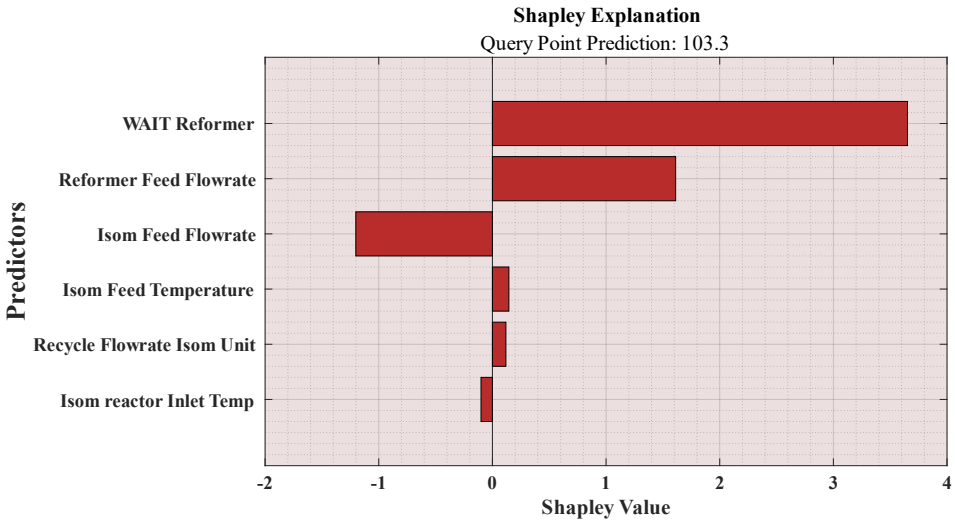

Figure 15 Shapley values of top six contributing predictor variables for RON prediction

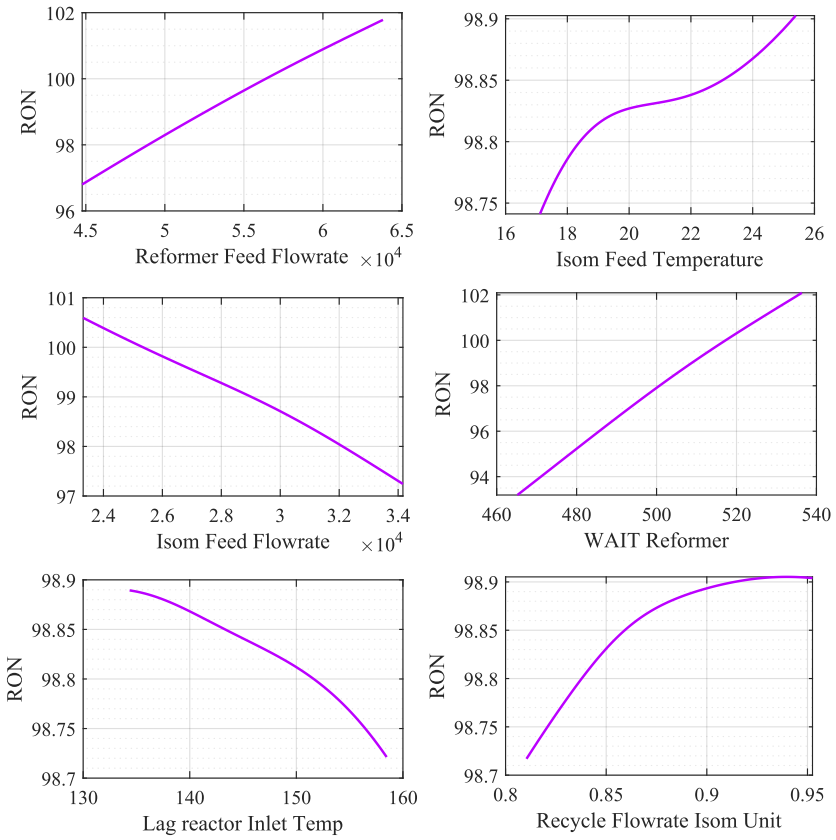

Figure 16 Partial Dependence Plots of predictors to RON prediction

Shapley analysis can give us the overview of the main contributing predictor variables, but to fully understand the trends between RON and predictor variables, partial dependency charts can give us better insights. Figure 16 shows the partial dependency charts of the five top contributing factors. The reformer's weighted average inlet temperature has the most impact on RON. This trend is in accordance with work related to RON dependence on reformer temperature [60]. Equilibrium calculations imply that hydrogen formation and aromatic reactions are favored at elevated temperatures. Furthermore, naphthenes and paraffin or consumed at a more significant rate at higher temperatures. Elevated temperature values higher than 520 ^{o}C result in aromatic forming reactions like hydroacylation reactions of paraffin to naphthene, paraffin conversion to isoparaffins, and naphthene dehydrogenation rates highly increased resulting in increased aromatic content which in turn increases RON [61].

4.1.4 SR comparison with GA and PSO based Optimization

Table 4 provides a comparison between straight run (SR), surrogate-based GA optimization, and particle swarm optimization (PSO). It is evident from the results that in all cases, GA and PSO optimize and increase the research octane number value as compared to the straight run conditions. In dataset 1, the SR value of 96.79 was increased to 101.72 and 101.90 by ANN-GA and ANN-PSO methodologies. The second observation that can be made from this table is that PSO-optimized values are higher as compared to the genetic algorithm in all cases for all surrogate models; however, this difference is not significantly high. The second section of the table contains the absolute error values of PSO and GA-based optimization of all the models. To check the closeness of optimized values to actual plant behavior, the results were validated by inserting the optimized parameters into the Aspen HYSYS software and then noting the RON values. Error value was calculated between the surrogate model-based optimization value and the Aspen HYSYS value. A comparison of error values indicates that ET and RT models show higher errors as compared to GPR, SVM, and ANN. This observation is in accordance with the prediction performance of the surrogate machine learning models. R^2 and RMSE values of GPR, SVM, and ANN are better compared to ET and RT, which results in superior fitness function

evaluations in evolutionary algorithms compared to ET and RT.

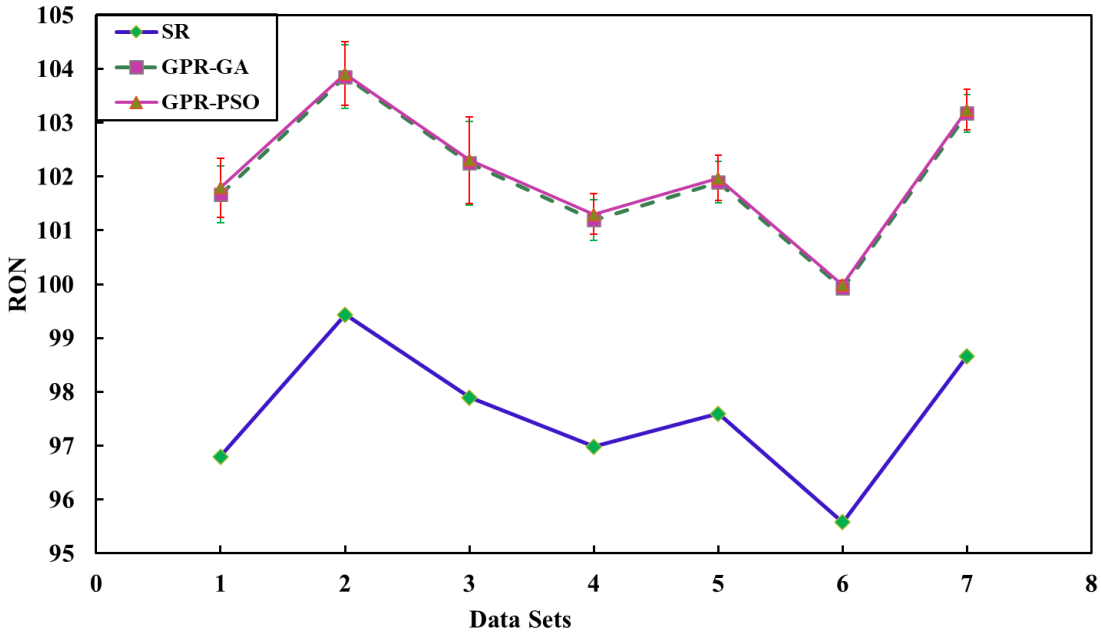

Figure 17 Comparison of straight run, GPR-GA, and GPR-PSO for RON prediction

Table 4 Comparison of SR, surrogate based GA optimization and PSO optimization

Datasets	**SR**	**GA**					**PSO**				
		GPR	**SVM**	**ANN**	**ET**	**RT**	**GPR**	**SVM**	**ANN**	**ET**	**RT**
1	96.795	101.662	101.656	101.726	99.643	98.927	101.792	101.849	101.902	99.643	98.927
2	95.579	99.926	99.978	100.670	100.928	100.074	99.982	100.022	101.594	100.952	100.074
3	98.665	103.177	99.957	99.966	101.104	100.074	103.236	100.022	100.031	101.122	100.074
4	98.498	102.990	102.989	102.876	101.861	101.454	103.039	103.076	102.943	101.877	101.634
5	96.943	101.315	101.430	101.231	101.514	101.454	101.358	101.522	101.333	101.514	101.454
6	98.666	103.136	103.198	102.842	101.899	101.454	103.236	103.230	103.004	101.919	101.454
Absolute Error Values											
1		0.216	0.546	0.436	0.572	0.309	0.416	1.349	1.432	0.457	1.203
2		0.046	0.090	0.227	0.252	0.825	1.640	1.419	1.066	1.324	1.399
3		0.351	0.378	0.229	0.249	0.095	0.110	1.402	1.386	0.629	0.692
4		0.437	0.262	0.445	0.463	0.124	0.130	0.610	0.853	0.356	1.001
5		0.158	0.147	0.449	0.479	0.051	0.072	0.389	0.325	2.148	1.195
6		0.364	0.364	0.539	0.555	0.126	0.136	0.599	0.672	0.739	0.123

4.1.5 Optimization Results of Integrated Naphtha Reforming and Isomerization Section

The global optimization toolbox of MATLAB 2022a was utilized to optimize the Integrated naphtha reforming and isomerization process by using the machine learning method as a surrogate model. Table 5 contains the parameters of GA and PSO that were employed to optimize the RON value. Optimized parameters of the integrated naphtha reforming and isomerization process are provided in table 6.

Table 5 Optimization parameters of GA and PSO for integrated naphtha reforming and isomerization process

GA parameters		**PSO parameters**	
Initial population	200	Swarm size	200
Crossover	Over scatter	Min Neighbors Fraction	0.25
Crossover probability	0.8	Self-Adjustment Weight	1.49
Mutation	Adapt feasible	Social Adjustment Weight	1.49
Selection	Tournament	Initial Swarm Span	2000

Table 6 GA and PSO based optimization of integrated naphtha reforming and isomerization reaction

Variables		**GPR-GA**	**GPR-PSO**
Flowrate	Isom Feed	28235.456	28224.500
	Reformer Feed	54511.873	54600.000
	Isom Unit Recycle	2996.544	3024.000
Temperature	Reformer Feed	20.483	20.745
	Isom Feed	20.941	21.000
	Product Heater	28.735	28.500
	WAIT Reformer	535.464	535.500
	Lead Reactor	127.893	128.100
	Lag reactor	138.129	137.750
H2HC Ratio	Reforming Unit	2.012	2.053
	Isom Unit	0.381	0.386
	RON	103.325	103.419

Table 7 Straight run conditions and sample datasets for integrated naphtha reforming and isomerization process

	Predictor Variables											**Response**
	Flowrates (kg/hr)			**Temperature (°C)**						**H_2HC-Ratio**		
Data sets	**Isom Feed**	**Reformer Feed**	**Recycle Isom Unit**	**Reformer Feed**	**Isom Feed**	**Product Heater**	**WAIT Reformer**	**Lead Reactor**	**Lag reactor**	**Reforming Unit**	**Isom Unit**	**RON**
SR	29708.25	52002.00	3081.90	20.00	20.00	30.00	510.00	122.00	145.00	2.00	0.37	99.14
1.00	30078.08	57730.24	3115.46	19.03	22.07	30.54	503.98	123.74	156.72	2.08	0.38	99.66
2.00	30568.27	53552.61	3146.26	20.92	19.89	30.43	500.15	122.47	146.19	2.06	0.36	98.06
3.00	30705.95	56605.22	3012.37	20.09	22.76	30.28	501.09	122.75	153.56	2.01	0.36	98.80
4.00	29708.31	51897.59	3113.46	20.81	20.22	31.08	517.79	122.47	142.70	2.02	0.36	100.04
5.00	27586.16	54225.77	3114.33	20.70	18.65	29.00	517.64	117.55	146.14	2.22	0.38	101.27
6.00	29314.43	54539.60	3011.47	19.08	20.77	30.89	499.77	120.55	143.81	2.08	0.38	98.58
7.00	28588.50	50470.51	3101.11	19.29	19.38	30.41	504.44	122.24	146.41	1.91	0.39	98.28
8.00	30367.05	51864.12	3157.32	19.31	19.68	29.08	516.08	120.43	148.03	2.02	0.35	99.55
9.00	30618.73	50987.52	3052.29	19.25	21.87	28.88	506.09	121.27	141.27	2.09	0.37	98.08
10.00	30270.72	57520.56	2969.73	21.14	21.37	30.33	531.70	123.74	139.32	2.19	0.35	102.56

4.2 Quality Prediction of Fluid catalytic Cracker

Table 13 contains the steady state values and sample datasets with the artificial uncertainties of ± 5 % in the input parameters of the fluid catalytic cracking unit.

4.2.1 Hyperparameter Optimization of Prediction Methods

Bayesian optimization was employed to optimize and select specific hyperparameters of each model by using the regression toolbox of MATLAB v2022a. 10-fold cross-validation was used to train and validate the performance of models for a specific set of hyperparameters. MSE was used as the selection criteria for hyperparameters, and optimization was performed, then, the hyperparameters giving the lowest MSE value were selected for further model development. Table 8 contains the hyperparameter search space along with optimized values for all models. Tuned hyperparameters of GPR were linear basis functions with isotropic Matern 5/2 kernel function and sigma value of 0.00102. SVM hyperparameters giving the best performance were the linear kernel function with a box constraint of 3.2310 and an epsilon value of 0.3231. ANN hyperparameters were 3 layers with 10 neurons each and ReLU as activation function. The leaf size of 20, the number of learners was 495, and the number of predictors to the sample value of 20 was hyperparameters of ensemble learning. LSBoost was used as the ensemble method.

Table 8 Hyperparameter tuning of ML methods for FCC

Methods	**Hyperparameters**	**Optimized Values**	**Hyperparameter Search Space**
GPR	Basis function Kernel function Sigma	Linear Isotropic Matern 5/2 0.00102	Constant, Zero, Linear Rational Quadratic Non/iso-tropic Exponential, Non/iso-tropic Matern 3/2, Non/iso-tropic Matern 5/2 0.0001-20.1131
SVM	Kernel function Box constraint Epsilon	Linear 3.2310 0.3231	Gaussian, Linear, Quadratic, Cubic 0.001-1000 0.0015295-152.9515
ET	Ensemble method Minimum leaf size	LSBoost 20	Bag, LSBoost 1-146

	Number of learners Number of predictors to sample	495 20	10-500 1-11
RT	Minimum leaf size	14	1-146
ANN	Layers Neurons in layers Activation Function In hidden layer	3 10 ReLU	1-3 1-100 ReLU, Sigmoid, Tanh, None

4.2.2 Prediction Performance Evaluation

The research octane number of high naphtha in FCC was predicted by training machine learning algorithms such as ANN, GPR, SVM, ET, and RT. Optimized hyperparameters were used to train the models. MSE, RMSE, and R^2 parameters were employed to compare the estimation capabilities of the above-mentioned methods. Values of training and testing RMSE, MSE, and R^2 are given in Table 9. Results point out that GPR, SVM, and ANN with R^2 values of 0.99637, 0.97925, 0.97865, MSE values of 0.01168, 0.06679, 0.06871, and RMSE values of 0.10807, 0.25845, 0.26212 respectively indicate high performance for prediction of RON.

The performance of regression and ensemble trees is on the lower end of the spectrum with R^2 values of 0.94160 and 0.89068, respectively. Regression and ensemble trees performed poorly and could not capture the nonlinearity associated with the RON prediction. RMSE of the regression tree and ensemble of trees was 0.59315 and 0.43355, respectively. The performance of the individual regression trees is lower as compared to the ensemble of trees. Bagging the trees increases the performance by averaging out the individual performance of the trees. Figure 18 shows the plots between actual and predicted values of RON for ANN, GPR, SVM, regression trees, and ensemble trees in terms of training and testing data. Scatter plots further solidify that GPR, SVM, and ANN are the better surrogate model for capturing the nonlinearity associated with the dependence of RON on the predictor variables.

Table 9 Prediction performance of ML methods for FCC

	R^2		MSE		RMSE	
Model	**Training**	**Testing**	**Training**	**Testing**	**Training**	**Testing**
GPR	0.9725	0.9701	0.1596	0.1527	0.3995	0.3908
SVM	0.9528	0.9697	0.2743	0.1547	0.5237	0.3934
ANN	0.9446	0.9632	0.3218	0.1878	0.5673	0.4333
ET	0.9360	0.9413	0.3719	0.2998	0.6098	0.5476
RT	0.8935	0.8802	0.6186	0.6117	0.7865	0.7821

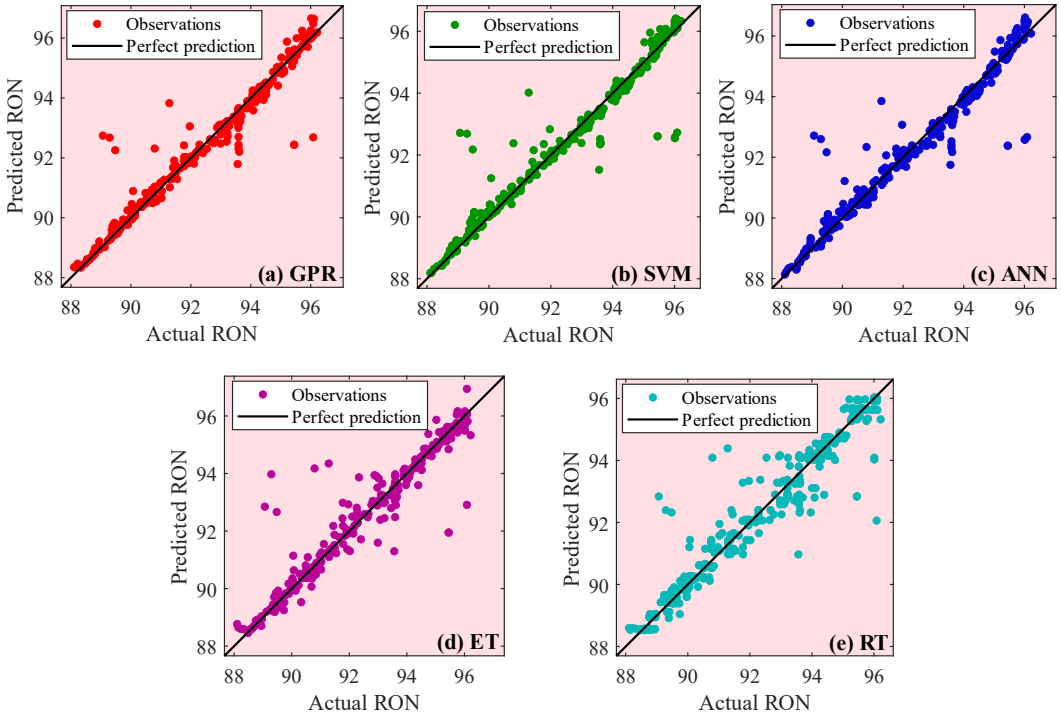

Figure 18 Prediction performance for (a) GPR, (b) SVM, (c) ANN, (d) ET and (e) RT

4.2.3 Partial Dependence Plots

To make machine learning models more intuitive and easier to comprehend SHapley Additive exPlanations (SHAP) tool was developed [62]. SHAP is based on the concept of shapley values which help in identifying the contribution of players and identifies the payout of the game accordingly. SHAP is used to comprehend the effect of different predictor variables on the output performance of machine learning models. As discussed in the previous section GPR outperformed as compared to other trained ML methods based on the R^2, RMSE, and MSE values. Hence for the GPR model, shapley explanation chart was created. Figure 19 contains the top 10 contributing

parameters in the prediction of RON. The reactor temperature of FCC has an obvious effect on the RON prediction as evident from Figure 19. Micro activity test (MAT) was found to be the second contributing factor and has a major effect on the RON prediction. Catalyst activity can have a tremendous effect on the quality of FCC products hence increased MAT means increased activity which results in better reaction rates, conversion, and selectivity which in turn increases RON [63].

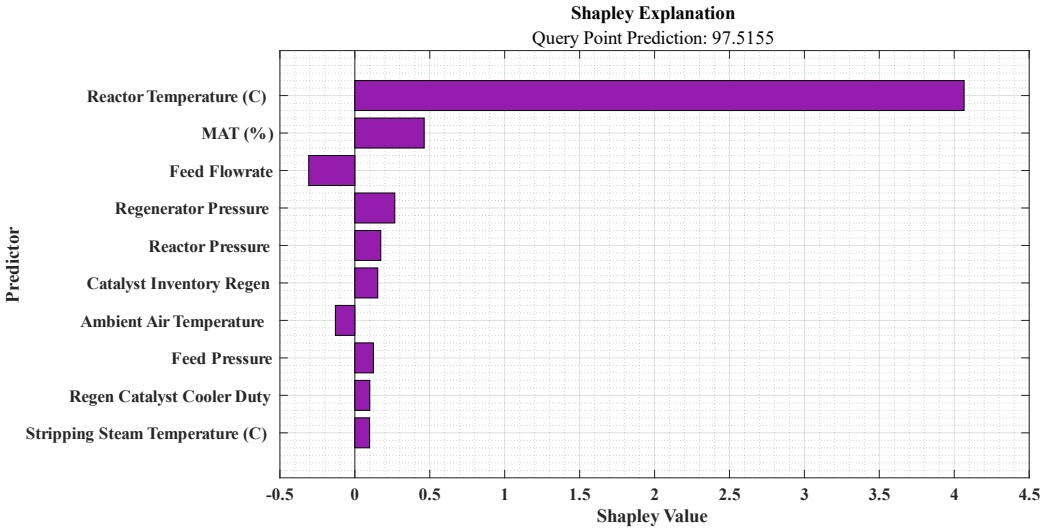

Figure 20 Shapley explanation chart of RON prediction in FCC

Shapley analysis identifies the contributing factors but to comprehend the dependence of prediction on input parameters, partial dependence charts provide a clearer picture. Figure 19 shows the partial dependency charts of the top contributing factors. As

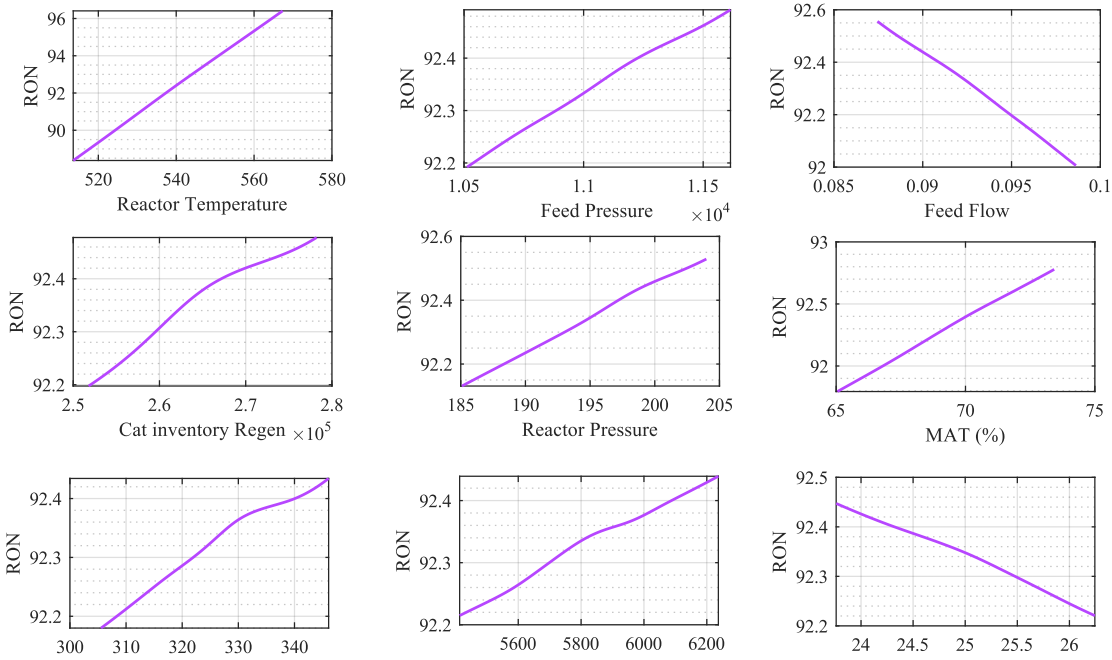

Figure 19 Partial dependance plots of top contributing factors in FCC

observed from shapley chart, reactor temperature and MAT have the most effect on the RON prediction. Increasing reaction temperature results in increasing reaction rates of cracking reactions resulting in enhancement of RON. PDP indicates that increasing the feed flow rate has a diverse effect on the RON of the product. This can be explained by low residence time due to an increase in feed flowrate which will lead to lower chances of reaction rate completion [64].

4.2.4 SR comparison with GA and PSO based Optimization

Table 10 provides a comparison between straight run (SR), surrogate-based GA optimization, and particle swarm optimization (PSO). It is evident from the results that in all cases, GA and PSO optimize and increase the research octane number value as compared to the straight run conditions. In dataset 1, the SR value of 92.7684 was increased to 98.336 and 98.576 by GPR-GA and GPR-PSO methodologies. The second observation that can be made from this table is that PSO-optimized values are higher as compared to the genetic algorithm in all cases; however, this difference is not significantly high. The table also contains the absolute error values of PSO and GA-based optimization of the GPR model. To check the closeness of optimized values to actual plant behavior, the results were validated by inserting the optimized parameters into the Aspen HYSYS software and then noting the RON values. Error value was calculated between the GPR model-based optimization value and the Aspen HYSYS value. Lower error values indicate that GPR-based GA and PSO optimization frameworks are capable of predicting and optimizing the RON value with high accuracy. PSO-based optimization results in a high percentage error as compared to GA-based optimization. Hence GPR-GA was found to be the best surrogate-based evolutionary optimization framework for RON prediction and optimization of the high naphtha product of FCC.

Table 10 Comparison of SR, GPR-GA and GPR-PSO frameworks

	SR	GPRGA	Hysys	Error (%)	GPRPSO	Hysys	Error (%)
1	92.7684	98.336	98.131	0.209	98.576	98.081	0.504
2	90.7066	96.834	96.657	0.184	97.041	96.674	0.379
3	90.3291	96.188	95.978	0.219	96.441	95.873	0.593
4	91.444	98.042	97.891	0.155	98.221	97.751	0.481
5	90.7554	96.830	96.415	0.430	97.013	96.579	0.449
6	90.5572	96.437	96.131	0.318	96.770	96.253	0.537
7	89.4775	95.602	95.245	0.375	95.770	95.000	0.811

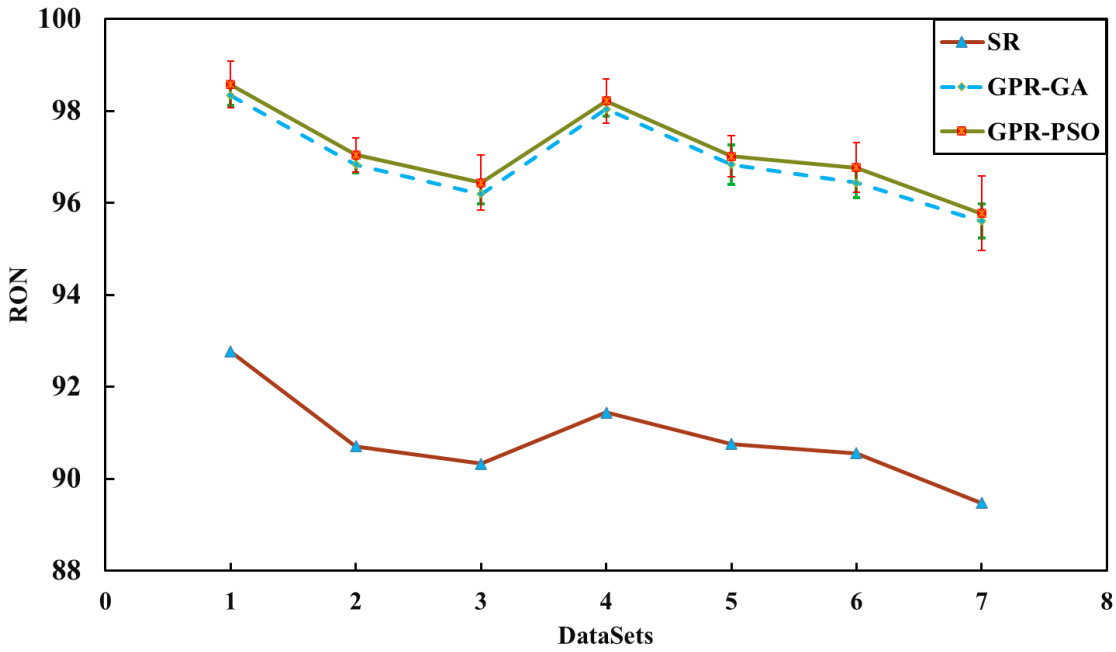

Figure 21 Straight run, GPR-PSO-GR-GA comparison

4.2.5 Optimization Results

GA and PSO are used to optimize the fluid catalytic process by using GPR as a surrogate model for fitness function evaluations. Table 11 contains the GA and PSO parameters that were used in optimization frameworks.

Table 11 GA and PSO parameters for RON optimization

GA parameters		**PSO parameters**	
Initial population	200	Swarm size	200
Crossover	Over scatter	Min Neighbors Fraction	0.25
Crossover probability	0.8	Self-Adjustment Weight	1.49
Mutation	Adapt feasible	Social Adjustment Weight	1.49
Selection	Tournament	Initial Swarm Span	2000

Table 12 contains the optimized values of input parameters of the fluid catalytic cracking unit.

Table 12 GPR based GA and PSO optimized parameters of FCC

Variables		**GPRGA**	**GPRPSO**
Temperature (^{o}C)	Feed	59.39	59.85
	Reactor	567.19	567.58
	Air Discharge	103.37	104.16
	Stripping Steam	345.34	346.29
	Dispersion Steam	345.87	346.29
	Ambient Air	24.00	23.75
Pressure	Feed	11461.04	11618.42
	Reactor	203.71	204.21
	Regenerator	143.40	143.46
	Stripping Steam	1424.10	1424.41
	Dispersion Steam	1315.32	1288.75
Flowrate	Feed	0.09	0.09

	Stripping Steam	1.09	1.09
	Dispersion Steam	1.54	1.55
Catalyst	Reactor Inventory	381348.36	380000.00
	Regenerator Inventory	278039.80	278250.00
	MAT	73.37	73.50
Heat Duty	Catalyst Cooler	6106.07	6154.49
HC Ratio	Coke	0.82	0.82
	RON HN	98.40	98.63

Table 13 Sample datasets for FCC

		Datasets						
		1	**2**	**3**	**4**	**5**	**6**	**7**
Temperature (ºC)	Feed	59.64	54.67	55.15	55.55	57.08	56.15	57.34
	Reactor	542.23	526.25	525.48	540.56	535.53	528.98	523.10
	Air Discharge	102.19	102.54	95.96	99.20	95.35	99.51	101.25
	Stripping Steam	327.56	315.34	330.58	329.80	339.36	325.86	342.94
	Dispersion Steam	318.90	321.12	313.55	323.26	318.18	313.70	340.29
	Ambient Air	24.95	25.87	26.14	25.00	25.45	26.09	24.47
Pressure	Feed	10639.85	11192.00	10665.32	10891.12	11490.38	11431.23	11456.98
	Reactor	190.68	195.91	202.29	195.39	187.77	195.46	193.44
	Regenerator	137.72	140.59	142.63	137.47	132.53	142.33	130.54
	Stripping Steam	1362.08	1365.99	1339.76	1356.58	1400.74	1391.00	1333.63
	Dispersion Steam	1378.52	1390.03	1392.17	1332.73	1355.90	1352.42	1392.38
Flowrate	Feed	0.09	0.09	0.09	0.09	0.09	0.09	0.09
	Stripping Steam	1.20	1.14	1.15	1.15	1.18	1.18	1.17
	Dispersion Steam	1.45	1.44	1.54	1.42	1.43	1.48	1.47
Catalyst	Reactor Inventory	401706.68	399851.56	390594.90	400000.00	392740.97	402712.87	384726.21
	Regenerator Inventory	252191.88	257613.54	263312.67	265000.00	254133.69	253179.68	266059.53
	MAT	68.45	73.20	71.26	66.58	66.88	68.86	67.20
Heat Duty	Catalyst Cooler	5651.88	6137.54	5583.65	5861.42	5648.24	6025.05	6154.20
HC Ratio	Coke	0.80	0.80	0.75	0.78	0.79	0.75	0.82
	RON HN	92.77	90.71	90.33	91.44	90.76	90.56	89.48

Conclusions

In this study, the effect of process variables like temperature, H_2HC ratios, flow rates, catalyst properties, pressures, utility temperature, utility pressure, ambient conditions, etc. on the research octane number was studied for two refinery processes, i.e. integrated naphtha isomerization and reforming process and fluid catalytic cracking. Uncertainty of ± 5\% was introduced in input data by using MATLAB and the Aspen HYSYS spreadsheet. Machine learning models such as GPR, ANN, SVM, ET, and RT were developed by using the uncertain inputs and the corresponding output values of RON. To enhance the performance of the machine learning methods, hyperparameter tuning was performed by utilizing Bayesian optimization in the regression learner app of MATLAB. Based on the performance of the machine learning models, the hierarchy of models for integrated naphtha reforming and isomerization process in descending order is as follows: GPR>ANN>SVM>ET>RT. The hierarchy of model performance in the case of fluid catalytic cracking in descending order is as follows: GPR>SVM >ANN >ET>RT. The top contributing factors in RON prediction were determined by using shapley values and partial dependence plots. The top machine learning model was then used as a surrogate model in PSO and GA optimization frameworks for fitness function evaluations. A comparison of optimization frameworks indicates that PSO performance is higher than GA.

References

[1]. J. D. Miller and C. Façanha, "The state of clean transport policy: a 2014 synthesis of vehicle and fuel policy developments," 2014.

[2]. E. ORTEGA. "An Overview of Hydrotreating." https://www.aiche.org/resources/publications/cep/2021/october/overview-hydrotreating (accessed.

[3]. M. A. Fahim, T. A. Al-Sahhaf, and A. Elkilani, *Fundamentals of petroleum refining*. Elsevier, 2009.

[4]. G. J. Antos and A. M. Aitani, *Catalytic naphtha reforming, revised and expanded*. CRC Press, 2004.

[5]. J. Ancheyta, A. Alvarez-Majmutov, and C. Leyva, "Hydrotreating of oil fractions," *Multiphase catalytic reactors: theory, design, manufacturing, and applications*, pp. 295-329, 2016.

[6]. G. Valavarasu and B. Sairam, "Light naphtha isomerization process: a review," *Petroleum Science and Technology*, vol. 31, no. 6, pp. 580-595, 2013.

[7]. G. Jiménez-García, R. Aguilar-López, and R. Maya-Yescas, "The fluidized-bed catalytic cracking unit building its future environment," *Fuel*, vol. 90, no. 12, pp. 3531-3541, 2011.

[8]. M. Ali, T. Tatsumi, and T. Masuda, "Development of heavy oil hydrocracking catalysts using amorphous silica-alumina and zeolites as catalyst supports," *Applied Catalysis A: General*, vol. 233, no. 1-2, pp. 77-90, 2002.

[9]. P. J. Ellis, C. A. Paul, and T. Session, *Tutorial: Delayed coking fundamentals*. American Institute of Chemical Engineers, 1998.

[10]. M. S. Rana, V. Sámano, J. Ancheyta, and J. Diaz, "A review of recent advances on process technologies for upgrading of heavy oils and residua," *Fuel*, vol. 86, no. 9, pp. 1216-1231, 2007.

[11]. C. Brock and D. L. Stanley, "The cooperative fuels research engine: applications for education and research," *Journal of Aviation Technology and Engineering*, vol. 2, no. 1, p. 7, 2012.

[12]. D. Özdemir, "Determination of octane number of gasoline using near infrared spectroscopy and genetic multivariate calibration methods," *Petroleum science and technology*, vol. 23, no. 9-10, pp. 1139-1152, 2005.

[13]. V. A. Rudnev, A. P. Boichenko, and P. V. Karnozhytskiy, "Classification of gasoline by octane number and light gas condensate fractions by origin with using dielectric or gas-chromatographic data and chemometrics tools," *Talanta,* vol. 84, no. 3, pp. 963-970, 2011.

[14]. J. Li, Y. Tan, and L. Liao, "Modeling and optimization of a semi-regenerative catalytic naphtha reformer," in *Proceedings of 2005 IEEE Conference on Control Applications, 2005. CCA 2005.*, 2005: IEEE, pp. 867-872.

[15]. M. Z. Stijepovic, A. Vojvodic-Ostojic, I. Milenkovic, and P. Linke, "Development of a kinetic model for catalytic reforming of naphtha and parameter estimation using industrial plant data," *Energy & Fuels,* vol. 23, no. 2, pp. 979-983, 2009.

[16]. I. Khosrozadeh, M. Talaghat, and A. Roosta, "Optimization of semiregenerative catalytic naphtha reforming unit to enhance octane number and reformate yield," *Iranian Journal of Chemical Engineering (IJChE),* vol. 15, no. 2, pp. 52-64, 2018.

[17]. K. Kavousi and N. Mokhtarian, "Simulation the Continuous Catalytic Reforming (Octanizer) Unit of Isfahan Refinery Gasoline Production Complex with PETROSIM Software," *International Science and Investigation journal,* vol. 4, no. 3, pp. 14-33, 2015.

[18]. N. V. Chekantsev, M. S. Gyngazova, and E. D. Ivanchina, "Mathematical modeling of light naphtha (C5, C6) isomerization process," *Chemical Engineering Journal,* vol. 238, pp. 120-128, 2014.

[19]. V. A. Chuzlov, N. V. Chekantsev, and E. D. Ivanchina, "Development of complex mathematical model of light naphtha isomerization and rectification processes," *Procedia Chemistry,* vol. 10, pp. 236-243, 2014.

[20]. V. A. Chuzlov, E. D. Ivanchina, D. Igor'M, and K. V. Molotov, "Simulation of light naphtha isomerization process," *Procedia Chemistry,* vol. 15, pp. 282-287, 2015.

[21]. V. A. Chuzlov, E. D. Ivanchina, N. V. Chekantsev, and K. V. Molotov, "Efficiency improvement of the light gasoline fractions isomerization by mathematical modeling," *Procedia Engineering,* vol. 113, pp. 131-137, 2015.

[22]. M. M. Said, T. S. Ahmed, and T. M. Moustafa, "Predictive modeling and optimization for an industrial Penex isomerization unit: A case study," *Energy & fuels,* vol. 28, no. 12, pp. 7726-7741, 2014.

[23]. J. Buitrago, D. Amaya, and O. Ramos, "Model and simulation of a pentane isomerization reactor for naphtha stream in oil Refining," *Contemporary Eng. Sci,* vol. 10, pp. 1255-1267, 2017.

[24]. A. M. Ahmed, A. T. Jarullah, F. M. Abed, and I. M. Mujtaba, "Modeling of an industrial naphtha isomerization reactor and development and assessment of a new isomerization process," *Chemical Engineering Research and Design,* vol. 137, pp. 33-46, 2018.

[25]. A. T. Jarullah, F. M. Abed, A. M. Ahmed, and I. M. Mujtaba, "Optimisation of several industrial and recently developed AJAM naphtha isomerization processes using model based techniques," *Computers & Chemical Engineering,* vol. 126, pp. 403-420, 2019.

[26]. M. Mohamed, W. Shehata, A. A. Halim, and F. Gad, "Improving gasoline quality produced from MIDOR light naphtha isomerization unit," *Egyptian Journal of Petroleum,* vol. 26, no. 1, pp. 111-124, 2017.

[27]. I. Ahmad, A. Ayub, M. Kano, and I. I. Cheema, "Gray-box soft sensors in process industry: Current practice, and future prospects in era of big data," *Processes,* vol. 8, no. 2, p. 243, 2020.

[28]. G. Zahedi, S. Mohammadzadeh, and G. Moradi, "Enhancing gasoline production in an industrial catalytic-reforming unit using artificial neural networks," *Energy & fuels,* vol. 22, no. 4, pp. 2671-2677, 2008.

[29]. S. Sadighi, R. S. Mohaddecy, and A. Norouzian, "Optimizing an industrial scale naphtha catalytic reforming plant using a hybrid artificial neural network and genetic algorithm technique," *Bulletin of Chemical Reaction Engineering & Catalysis,* vol. 10, no. 2, p. 210, 2015.

[30]. F. Elfghi, "A hybrid statistical approach for modeling and optimization of RON: A comparative study and combined application of response surface methodology (RSM) and artificial neural network (ANN) based on design of experiment (DOE)," *Chemical Engineering Research and Design,* vol. 113, pp. 264-272, 2016.

[31]. S. Sadighi and R. S. Mohaddecy, "Predictive modeling for an industrial naphtha reforming plant using artificial neural network with recurrent layers," *International Journal of Technology,* vol. 2, pp. 1-11, 2013.

[32]. A. Al-Shathr, Z. M. Shakor, H. S. Majdi, A. A. AbdulRazak, and T. M. Albayati, "Comparison between Artificial Neural Network and Rigorous

Mathematical Model in Simulation of Industrial Heavy Naphtha Reforming Process," *Catalysts,* vol. 11, no. 9, p. 1034, 2021.

[33]. I. Ahmad, G. Ali, M. Bilal, A. Chughtai, A. Hussain, and M. Kano, "Quantitative analysis of product quality of naphtha reforming process under uncertain process conditions," *Chemical Engineering Communications,* vol. 207, no. 8, pp. 1092-1102, 2020.

[34]. S. Herceg, Ž. U. Andrijić, and N. Bolf, "Development of soft sensors for isomerization process based on support vector machine regression and dynamic polynomial models," *Chemical Engineering Research and Design,* vol. 149, pp. 95-103, 2019.

[35]. K. Toch, J. Thybaut, B. Vandegehuchte, C. Narasimhan, L. Domokos, and G. Marin, "A Single-Event MicroKinetic model for “ethylbenzene dealkylation/xylene isomerization” on Pt/H-ZSM-5 zeolite catalyst," *Applied Catalysis A: General,* vol. 425, pp. 130-144, 2012.

[36]. S. Sadighi, S. Zahedi, R. Hayati, and M. Bayat, "Studying Catalyst Activity in an Isomerization Plant to Upgrade the Octane Number of Gasoline by Using a Hybrid Artificial-Neural-Network Model," *Energy Technology,* vol. 1, no. 12, pp. 743-750, 2013.

[37]. O. Grozdanić, K. Sertić-Bionda, and M. Mužic, "The influence of process parameters on light naphtha isomerization," *Međunarodni znanstveno-stručni skup, 14. Ružičkini dani," Danas znanost-sutra industrija", Vukovar, Croatia, 13.-15. rujna, 2012,* pp. 108-120, 2013.

[38]. M. Wei, M. Yang, F. Qian, and W. Du, "Optimization of catalytic naphtha reforming process based on modified differential evolution algorithm," *IFAC-PapersOnLine,* vol. 48, no. 8, pp. 373-378, 2015.

[39]. R. Zainullin, A. Zagoruiko, K. Koledina, I. Gubaidullin, and R. Faskhutdinova, "Multi-criterion optimization of a catalytic reforming reactor unit using a genetic algorithm," *Catalysis in Industry,* vol. 12, no. 2, pp. 133-140, 2020.

[40]. B. S. Babaqi, M. S. Takriff, N. T. A. Othman, and S. K. Kamarudin, "Yield and energy optimization of the continuous catalytic regeneration reforming process based particle swarm optimization," *Energy,* vol. 206, p. 118098, 2020.

[41]. P. Pasandide and M. Rahmani, "Simulation and optimization of continuous catalytic reforming: Reducing energy cost and coke formation," *International Journal of Hydrogen Energy*, vol. 46, no. 58, pp. 30005-30018, 2021.

[42]. L. Mencarelli, Q. Chen, A. Pagot, and I. E. Grossmann, "A review on superstructure optimization approaches in process system engineering," *Computers & Chemical Engineering,* vol. 136, p. 106808, 2020.

[43]. P. Duchêne, L. Mencarelli, and A. Pagot, "Optimization approaches to the integrated system of catalytic reforming and isomerization processes in petroleum refinery," *Computers & Chemical Engineering,* vol. 141, p. 107009, 2020.

[44]. B. S. Babaqi, M. S. Takriff, N. T. A. Othman, H. A. Hasan, and E. Mahmoudi, "Modeling and Simulation for Integrating Four Reactors in the Continuous Catalytic Regeneration Reforming Process Using MATLAB," in *Journal of Physics: Conference Series,* 2021, vol. 1962, no. 1: IOP Publishing, p. 012002.

[45]. T. Sharma, A. Sharma, S. Verma, and S. Mohile, "Sensitivity analysis and grouping of rate constants of aspen kinetic model for a fluidized catalytic cracker," in *AIP Conference Proceedings*, 2021, vol. 2412, no. 1: AIP Publishing LLC, p. 050003.

[46]. Y. Zhang, Z. Li, Z. Wang, and Q. Jin, "Optimization Study on Increasing Yield and Capacity of Fluid Catalytic Cracking (FCC) Units," *Processes,* vol. 9, no. 9, p. 1497, 2021.

[47]. D. Ahmed and S. Ateya, "Modelling and Simulation of Fluid Catalytic Cracking Unit," *J. Chem. Eng. Process Technol,* vol. 7, no. 4, p. 308, 2016.

[48]. Y. Zhang *et al.*, "Modeling, simulation, and optimization for producing ultra-low sulfur and high-octane number gasoline by separation and conversion of fluid catalytic cracking naphtha," *Fuel,* vol. 299, p. 120740, 2021.

[49]. W. Tian, S. Wang, S. Sun, C. Li, and Y. Lin, "Intelligent prediction and early warning of abnormal conditions for fluid catalytic cracking process," *Chemical Engineering Research and Design,* vol. 181, pp. 304-320, 2022.

[50]. C. Chen, L. Zhou, X. Ji, G. He, Y. Dai, and Y. Dang, "Adaptive modeling strategy integrating feature selection and random forest for fluid catalytic cracking processes," *Industrial & Engineering Chemistry Research,* vol. 59, no. 24, pp. 11265-11274, 2020.

[51]. O. Santander, V. Kuppuraj, C. A. Harrison, and M. Baldea, "Integrated deep learning-production planning-economic model predictive control framework for large-scale processes. A fluid catalytic cracker-fractionator case study," *Computers & Chemical Engineering,* vol. 167, p. 107977, 2022.

[52]. F. Yang, C. Dai, J. Tang, J. Xuan, and J. Cao, "A hybrid deep learning and mechanistic kinetics model for the prediction of fluid catalytic cracking performance," *Chemical Engineering Research and Design,* vol. 155, pp. 202-210, 2020.

[53]. J. Long, T. Li, M. Yang, G. Hu, and W. Zhong, "Hybrid strategy integrating variable selection and a neural network for fluid catalytic cracking modeling," *Industrial & Engineering Chemistry Research,* vol. 58, no. 1, pp. 247-258, 2018.

[54]. N. A. M. Ahmed and M. A. Mustafa, "Developing Regression Models for the Characterization of Spent Fluid Catalytic Cracking Catalyst."

[55]. C. Chen, N. Lu, L. Wang, and Y. Xing, "Intelligent selection and optimization method of feature variables in fluid catalytic cracking gasoline refining process," *Computers & Chemical Engineering,* vol. 150, p. 107336, 2021.

[56]. X. Chen *et al.*, "Real-time refinery optimization with reduced-order fluidized catalytic cracker model and surrogate-based trust region filter method," *Computers & Chemical Engineering,* vol. 153, p. 107455, 2021.

[57]. I.-S. Han, J. B. Riggs, and C.-B. Chung, "Multivariable control of a fluidized catalytic cracking process under full and partial combustion modes," *Journal of chemical engineering of Japan,* vol. 35, no. 9, pp. 830-839, 2002.

[58]. T. Dias, R. Oliveira, P. Saraiva, and M. S. Reis, "Predictive analytics in the petrochemical industry: Research Octane Number (RON) forecasting and analysis in an industrial catalytic reforming unit," *Computers & Chemical Engineering,* vol. 139, p. 106912, 2020.

[59]. M. R. Rahimpour, M. Jafari, and D. Iranshahi, "Progress in catalytic naphtha reforming process: A review," *Applied energy,* vol. 109, pp. 79-93, 2013.

[60]. U. T. Turaga and R. Ramanathan, "Catalytic naphtha reforming: revisiting its importance in the modern refinery," 2003.

[61]. J. Martínez, M. A. Zúñiga-Hinojosa, and R. S. Ruiz-Martínez, "A Thermodynamic Analysis of Naphtha Catalytic Reforming Reactions to Produce High-Octane Gasoline," *Processes,* vol. 10, no. 2, p. 313, 2022.

[62]. E. Stenwig, G. Salvi, P. S. Rossi, and N. K. Skjærvold, "Comparative analysis of explainable machine learning prediction models for hospital mortality," *BMC Medical Research Methodology,* vol. 22, no. 1, pp. 1-14, 2022.

[63]. B. Siddiqui, A. Aitani, M. Saeed, and S. Al-Khattaf, "Enhancing the production of light olefins by catalytic cracking of FCC naphtha over mesoporous ZSM-5 catalyst," *Topics in Catalysis,* vol. 53, no. 19, pp. 1387-1393, 2010.

[64]. S. Zahedi Abghari, "Sensitivity Analysis and Development of a Set of Rules to Operate FCC Process by Application of a Hybrid Model of ANFIS and Firefly Algorithm," *Journal of Petroleum Science and Technology,* vol. 9, no. 3, pp. 10-26, 2019.