



**INSTITUTO LATINO-AMERICANO DE
TECNOLOGIA, INFRAESTRUTURA E TERRITÓRIO
(ILATIT)**

ENGENHARIA QUÍMICA

APLICAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA PARA A MODELAGEM DE UMA COLUNA DE DESTILAÇÃO EXTRATIVA

ÁLVARO PALMA DA SILVA

Foz do Iguaçu

2022



**INSTITUTO LATINO-AMERICANO DE
TECNOLOGIA, INFRAESTRUTURA E TERRITÓRIO
(ILATIT)**

ENGENHARIA QUÍMICA

**APLICAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA
PARA A MODELAGEM DE UMA COLUNA DE DESTILAÇÃO
EXTRATIVA**

ÁLVARO PALMA DA SILVA

Trabalho de Conclusão de Curso apresentado ao Instituto Latino-Americano de Tecnologia, Infraestrutura e Território da Universidade Federal da Integração Latino-Americana, como requisito parcial à obtenção do título de Bacharel em Engenharia Química

Orientador: Prof^o. Dr. Luís Antonio Lourenço

Foz do Iguaçu

2022

ÁLVARO PALMA DA SILVA

**APLICAÇÃO DE MODELOS DE APRENDIZADO DE MÁQUINA
PARA A MODELAGEM DE UMA COLUNA DE DESTILAÇÃO
EXTRATIVA**

Trabalho de Conclusão de Curso apresentado ao Instituto Latino-Americano de Tecnologia, Infraestrutura e Território da Universidade Federal da Integração Latino-Americana, como requisito parcial à obtenção do título de Bacharel em Engenharia Química

BANCA EXAMINADORA

Orientador: Prof^o. Dr. Luís Antonio Lourenço
UNILA

Prof. Convidado 1
UNILA

Prof. Convidado 2
UNILA

Foz do Iguaçu, ___ de _____ de ____.

TERMO DE SUBMISSÃO DE TRABALHOS ACADÊMICOS

Nome completo do autor(a): _____

Curso: _____

Tipo de Documento

- (.....)graduação (.....)artigo
(.....)especialização (.....)trabalho de conclusão de curso
(.....)mestrado (.....)monografia
(.....)doutorado (.....)dissertação
(.....)tese
(.....)CD/DVD - obras audio/visuais
(.....)_____

Título do trabalho acadêmico: _____

Nome do orientador: _____

Data da Defesa: ____/____/_____

Licença não-exclusiva de Distribuição

O referido autor(a):

a) Declara que o documento entregue é seu trabalho original, e que o detém o direito de conceder os direitos contidos nesta licença. Declara também que a entrega do documento não infringe, tanto quanto lhe é possível saber, os direitos de qualquer outra pessoa ou entidade.

b) Se o documento entregue contém material do qual não detém os direitos de autor, declara que obteve autorização do detentor dos direitos de autor para conceder à UNILA – Universidade Federal da Integração Latino-Americana os direitos requeridos por esta licença, e que esse material cujos direitos são de terceiros está claramente identificado e reconhecido no texto ou conteúdo do documento entregue.

Se o documento entregue é baseado em trabalho financiado ou apoiado por outra instituição que não a Universidade Federal da Integração Latino-Americana, declara que cumpriu quaisquer obrigações exigidas pelo respectivo contrato ou acordo.

Na qualidade de titular dos direitos do conteúdo supracitado, o autor autoriza a Biblioteca Latino-Americana – BIUNILA a disponibilizar a obra, gratuitamente e de acordo com a licença pública Creative Commons **Licença 3.0 Unported**.

Foz do Iguaçu, _____ de _____ de _____.

Assinatura do Responsável

AGRADECIMENTOS

Os agradecimentos principais são a Deus e a minha família, pelo apoio do começo ao fim desse desafio.

Em especial a minha mãe Denise Cristina Teixeira Palma, por não hesitar em me apoiar quando eu disse que largaria tudo pra conseguir essa graduação.

A minha parceira de vida Bruna Baldasso, pelo apoio incondicional e por me inspirar a seguir evoluindo sempre.

Ao meu orientador Prof^o. Dr. Luís Antonio Lourenço por compartilhar seus estudos e me direcionar sempre com empatia e objetividade.

Ao Prof. Dr. Raphael Fortes pela amizade valiosa e a honra de travar uma grande batalha ao seu lado durante essa etapa.

A Prof. Dra. Andréia Cristina Furtado por ser uma líder atenta e compreensível para os integrantes desse curso.

Ao Prof. Dr. Cleilton Aparecido Canal por ter sido capaz de transmitir os conhecimentos de geometria analítica, cálculo e reforçar a segurança de que posso aprender qualquer coisa.

A escola Trybe, por ensinar ferramentas que usei aqui e que com certeza usarei a partir daqui.

Aos meus amigos, pela fidelidade e por compreenderem minhas ausências ao longo do curso.

Aos meus colegas, pelas experiências, pela união e pelas risadas.

Aos demais professores do curso de Engenharia Química, pelos ensinamentos valiosos.

À Universidade Federal da Integração Latino-Americana, o corpo docente, técnicos administrativos e de laboratório, e aos demais terceirizados pelo trabalho e dedicação, contribuindo com essa formação.

*“Somos os Cavaleiros. Ficamos entre
a luz e a escuridão e mantemos seu equilíbrio.
Ignorância, medo, ódio: esses são os nossos inimigos.”
(Eldest, Christopher Paolini)*

RESUMO

Os processos químicos industriais são essenciais ao modo de vida contemporâneo, pois os produtos gerados pela indústria química abastecem tanto outras indústrias, quanto o público em geral. A atuação do engenheiro químico pode ter grande impacto social, econômico e ambiental, o que traz uma responsabilidade de ser manter atualizado as condições do ambiente que exerce influência. Por essa razão, este trabalho busca associar ferramentas de inteligência artificial para modelagem de processos químicos industriais. O processo escolhido para esse estudo foi a extração de metanol, que é produzido a partir do gás natural (metano) e pode ser purificado com o glicerol como solvente. O Brasil dispõe tanto de gás natural, quanto de glicerol na faixa de milhões de toneladas por ano, pois é um dos maiores produtores de biodiesel do mundo e o glicerol é um subproduto gerado em grande quantidade na produção do biodiesel. Para purificar o metanol é empregada uma operação unitária de separação composta por uma ou mais colunas de destilação extrativas. Este equipamento é tradicionalmente projetado pelo engenheiro químico usando modelos baseados nos princípios físicos de conservação da massa e energia. São modelos complexos que podem demandar muito tempo para desenvolvimento em processos específicos. Por outro lado, existem modelos baseados nos conceitos de inteligência artificial e ciência de dados que são mais simples e podem se adaptar rapidamente ao processo. Neste trabalho foram aplicados os modelos supervisionados de aprendizado de máquina floresta aleatória e regressão linear para a modelagem da coluna de destilação extrativa do metanol usando glicerol como solvente. O modelo floresta aleatória apresentou boa aderência aos dados e realizou previsões próximas dos valores esperados. O modelo de regressão linear apresentou dificuldade de aderência aos dados em algumas faixas de valores mais baixos, pois as interações do processo a variações em determinadas entradas possuem características não lineares.

Palavras-chave: árvore de decisão. inteligência artificial. ciência de dados.

RESUMEN

Los procesos químicos industriales son esenciales para la forma de vida contemporánea, ya que los productos generados por la industria química abastecen tanto a otras industrias como al público en general. La actuación del ingeniero químico puede tener un gran impacto social, económico y ambiental, lo que trae consigo la responsabilidad de mantener actualizadas las condiciones del medio que lo influencia. Por eso, este trabajo busca asociar herramientas de inteligencia artificial para la modelización de procesos químicos industriales. El proceso escogido para este estudio fue la extracción de metanol, el cual se produce a partir del gas natural (metano) y puede ser purificado con glicerol como solvente. Brasil tiene tanto gas natural como glicerol en el rango de millones de toneladas por año, ya que es uno de los mayores productores de biodiesel del mundo y el glicerol es un subproducto generado en grandes cantidades en la producción de biodiesel. Para purificar el metanol se utiliza una operación de separación unitaria que comprende una o más columnas de destilación extractiva. Este equipo es diseñado tradicionalmente por el ingeniero químico utilizando modelos basados en los principios físicos de conservación de la masa y la energía. Estos son modelos complejos que pueden tomar mucho tiempo para desarrollarse en procesos específicos. Por otro lado, existen modelos basados en los conceptos de inteligencia artificial y ciencia de datos que son más simples y pueden adaptarse rápidamente al proceso. En este trabajo, se aplicaron modelos de aprendizaje automático de bosque aleatorio supervisado y regresión lineal para modelar la columna de destilación extractiva de metanol utilizando glicerol como solvente. El modelo de bosque aleatorio mostró una buena adherencia a los datos y realizó predicciones cercanas a los valores esperados. El modelo de regresión lineal mostró dificultad en la adherencia a los datos en algunos rangos de valores inferiores, ya que las interacciones del proceso ante variaciones en ciertos insumos tienen características no lineales.

Palabras clave: árbol de decisión. inteligencia artificial. ciencia de los datos.

ABSTRACT

Industrial chemical processes are essential to the contemporary way of life, as the products generated by the chemical industry supply both other industries and the general public. The performance of the chemical engineer can have great social, economic and environmental impact, which brings a responsibility to keep up to date the conditions of the environment that influences it. For this reason, this work seeks to associate artificial intelligence tools for modeling industrial chemical processes. The process chosen for this study was the extraction of methanol, which is produced from natural gas (methane) and can be purified with glycerol as a solvent. Brazil has both natural gas and glycerol in the range of millions of tons per year, as it is one of the largest biodiesel producers in the world and glycerol is a by-product generated in large quantities in the production of biodiesel. In order to purify the methanol, a unit separation operation comprising one or more extractive distillation columns is used. This equipment is traditionally designed by the chemical engineer using models based on the physical principles of conservation of mass and energy. These are complex models that can take a long time to develop in specific processes. On the other hand, there are models based on the concepts of artificial intelligence and data science that are simpler and can quickly adapt to the process. In this work, supervised random forest machine learning and linear regression models were applied to model the extractive distillation column of methanol using glycerol as solvent. The random forest model showed good adherence to the data and made predictions close to the expected values. The linear regression model showed difficulty in adherence to the data in some ranges of lower values, as the process interactions to variations in certain inputs have non-linear characteristics.

Keywords: decision tree. artificial intelligence. data science.

LISTA DE ILUSTRAÇÕES

Figura 1 – Principais setores industriais de utilização da glicerina	18
Figura 2 – Representação de uma coluna de destilação	19
Figura 3 – Representação do funcionamento do algoritmo de árvores de decisão	23
Figura 4 – Comparação entre árvore de decisão e floresta de decisão	23
Figura 5 – Diagrama de divisão de tarefas do método	27
Figura 6 – Etapas realizadas pelo código de simulação	28
Figura 7 – Identificação das variáveis de funcionalidade e alvo no <i>Orange</i>	30
Figura 8 – Esquema de trabalho completo no <i>Orange</i>	31
Figura 9 – Valores estatísticos dos dados simulados.	32
Figura 10 – Valores da correlação linear de Pearson.	33
Figura 11 – Valores de comparação do ajuste preliminar de cada modelo de aprendizado de máquina.	34
Figura 12 – Tendência das médias nas árvores de decisões geradas.	35
Figura 13 – Aproximação nos primeiros nós	35
Figura 14 – Valores de predição do resultado de cada modelo de aprendizado de máquina.	36
Figura 15 – Gráfico de dispersão do modelo de regressão linear	37
Figura 16 – Gráfico de dispersão do modelo de floresta aleatória	38

LISTA DE ABREVIATURAS E SIGLAS

RR	Razão de Refluxo
BioSTEAM	The Biorefinery Simulation and Techno-Economic Analysis Modules
MSE	Erro Quadrático Médio
RMSE	Raíz Quadrada do Erro Médio
MAE	Erro Absoluto Médio
R ²	Coefficiente de Determinação

SUMÁRIO

1	INTRODUÇÃO	13
2	JUSTIFICATIVA	14
3	OBJETIVOS	15
3.1	OBJETIVO GERAL	15
3.2	OBJETIVOS ESPECÍFICOS	15
4	REVISÃO BIBLIOGRÁFICA	16
4.1	PROCESSOS QUÍMICOS INDUSTRIAIS	16
4.2	PROCESSOS DE PRODUÇÃO DO BIODIESEL E DO METANOL	17
4.3	COLUNA DE DESTILAÇÃO	19
4.4	MODELAGEM DE PROCESSOS INDUSTRIAIS	19
4.4.1	Projeto de uma Coluna de Destilação	20
4.5	MODELOS DE APRENDIZADO DE MÁQUINA	21
4.6	CIÊNCIA DE DADOS NA ENGENHARIA QUÍMICA	24
5	MATERIAIS E MÉTODOS	26
5.1	FERRAMENTAS UTILIZADAS	26
5.1.1	Ferramentas para Obtenção dos Dados	26
5.1.2	Ferramentas de Modelagem por Aprendizado de Máquina	26
5.2	MÉTODO DE MODELAGEM POR APRENDIZADO DE MÁQUINA	26
5.2.1	Diagrama do Método Utilizado	26
5.2.2	Etapa de Simulação	27
5.2.3	Etapa de Modelagem por Aprendizado de Máquina	29
6	RESULTADOS E DISCUSSÃO	32
6.1	ANÁLISE EXPLORATÓRIA DOS DADOS	32
6.2	AJUSTE DOS MODELOS SELECIONADOS	33
6.3	ÁRVORES DE DECISÃO GERADAS	34
6.4	PREDIÇÕES DO MODELO	36
7	CONSIDERAÇÕES FINAIS	39
	Referências	40
	Apêndices	45
APÊNDICE A	Código de simulação da coluna de destilação com módulo BioSTEAM	46
APÊNDICE B	Código com funções de inserção de dados e registro em arquivo CSV	47
APÊNDICE C	Código do arquivo de dados CSV	49
APÊNDICE D	Histogramas de distribuição dos dados simulados	50

Anexos	57
ANEXO A Simulador BioSTEAM: Módulos de Simulação de Biorrefinarias e Análises Tecno-econômicas	58
ANEXO B Ferramentas para inserção de dados no simulador, registro de resultados obtidos e geração de arquivo de trabalho	59
ANEXO C Software Orange Data Mining para avaliação, treinamento e testes de modelos de aprendizado de máquina	60

1 INTRODUÇÃO

Quando se planeja analisar e otimizar processos químicos industriais, modelos baseados nas equações fundamentais da física são a escolha tradicional, uma vez que estão bem definidos e confirmados pela literatura. Porém, são bastante complexos e demandam conhecimento dos fenômenos associados a operação industrial, o que nem sempre é possível. Por outro lado, os modelos de aprendizado de máquina são conhecidos como modelos de “caixa preta”, uma vez que não são baseados nos princípios fundamentais. Isto os tornam mais fáceis de serem construídos em relação à modelagem de processos com base nos fenômenos de transferência, na termodinâmica e cinética química. Entretanto, os modelos de aprendizagem de máquina necessitam de um grande volume de dados para treinar o algoritmo (BIKMUKHAMETOV; JÄSCHKE, 2020).

Kiss (2011) destaca que o uso da ciência de dados na indústria não é novidade em áreas como predição do efeito de perturbações em um sistema de produção, no entanto, na área de modelagem e otimização de processos ainda existem muitos campos a se estudar.

Devido aos avanços no processo de digitalização da indústria de manufatura e aos dados disponíveis resultantes, há um enorme progresso e grande interesse na integração de aprendizado de máquina e métodos de otimização no chão de fábrica para melhorar os processos de produção (WEICHERT et al., 2019).

Neste cenário, o objetivo deste trabalho é construir um modelo de aprendizado de máquina para análise e otimização do processo de destilação extrativa do metanol. Para isto, serão utilizados bancos de dados de uma coluna de destilação simulada.

2 JUSTIFICATIVA

O Brasil está atualmente entre os maiores produtores de biodiesel do mundo. No entanto, o processo de produção deste biocombustível gera como subproduto o glicerol em quantidades de milhões de toneladas por ano que tem se tornado um tópico urgente do ponto de vista econômico, ambiental e social (QUISPE; CORONADO; JR, 2013). Como ponto positivo, este subproduto ao ser utilizado como solvente apresenta boa compatibilidade com compostos orgânicos, alto ponto de ebulição e baixa agressividade ao meio ambiente (GU; JÉRÔME, 2010) e (KISS; IGNAT, 2012).

Além disso, existe uma disponibilidade crescente de gás natural no Brasil. Estimativas da Empresa Brasileira de Pesquisa Energética (EPE), dizem que até 2029, a disponibilidade de gás natural deve mais do que duplicar, atingindo os 138 milhões de metros cúbicos por dia. Sem novos investimentos em infraestrutura, há o risco de que uma fração importante do gás natural disponível não seja aproveitada (LEÃO, 2020).

O gás natural é a matéria prima para a produção do metanol, que é utilizado como combustível, solvente industrial e para diversas outras utilidades. Para que se produza um metanol de qualidade é preciso que a planta de produção seja desenvolvida com uma etapa de purificação. Esta purificação se trata de uma extração do metanol de uma mistura com água, e para que o ponto de azeotropia seja quebrado pode ser aplicado o glicerol como solvente (KISS; IGNAT, 2012). A operação unitária de separação empregada nessa extração é composta essencialmente por uma coluna de destilação.

Os métodos tradicionais de modelagem de uma coluna de destilação utilizam-se das leis de conservação fundamentais da física e possuem uma alta complexidade de desenvolvimento em relação a modelos mais novos que se baseiam em conceitos de inteligência artificial e ciência de dados.

Este trabalho busca empregar métodos de aprendizado de máquina, um subcampo da inteligência artificial e da ciência de dados, para modelar uma coluna de destilação de extração de metanol utilizando glicerol como solvente, buscando uma solução mais simples, rápida, adaptável e mutável que os modelos tradicionais (DOGAN; BIRANT, 2021), gerando assim, uma alternativa para futuras otimizações, predições e tomadas de decisão sobre este processo.

3 OBJETIVOS

3.1 OBJETIVO GERAL

O objetivo deste trabalho é utilizar métodos de aprendizado de máquina para desenvolver um modelo teórico computacional de uma coluna de destilação extrativa de metanol utilizando glicerol como solvente.

3.2 OBJETIVOS ESPECÍFICOS

- Simular o processo de extração do metanol e coletar os dados gerados;
- Preparar os dados obtidos e utilizá-los no treinamento de algoritmos de aprendizado de máquina;
- Obter um modelo computacional capaz de realizar previsões sobre o processo.

4 REVISÃO BIBLIOGRÁFICA

4.1 PROCESSOS QUÍMICOS INDUSTRIAIS

Uma indústria ou planta de produção tem suas características de acordo com o tipo de processo químico que é empregado em seu sistema.

Um processo químico é descrito como o processo em quem um composto químico é transformado em outro. A multibilionária indústria de processos químicos fabrica, processa, e produz vários tipos de compostos, substâncias químicas e produtos. Estes incluem petroquímicos e compostos inorgânicos, fármacos, gases industriais, plásticos, borrachas, detergentes, tintas, seladores, pigmentos e muitos outros. É uma indústria essencial para a economia global moderna e desempenha um papel vital na existência de quase toda indústria (INSPECTIONEERING, 2021).

O desenvolvimento de um novo processo em uma indústria desde a avaliação do conceito até uma realidade lucrativa é frequentemente um problema de enorme complexidade (AERSTIN; STREET, 2012). Um processo químico pode conter operações de diversos tipos, desde a combinação de soluções, separação de componentes até reação química. Em Engenharia Química estes processos são definidos como operações unitárias (LITTLE; WALKER; LEWIS, 1880).

A importância dos processos químicos industriais vai além do produto gerado. Turton et al. (2008), em seu livro comenta sobre a complexidade de processos químicos industriais e suas consequências sérias em caso de erros que incluem explosões, dano ao meio ambiente e riscos a saúde das pessoas, pois frequentemente estes processos envolvem substâncias químicas de alta reatividade, alta toxicidade, corrosividade e altas temperaturas e pressões. Ele enfatiza como é essencial que não ocorram erros ou omissões entre os envolvidos no desenvolvimento e operação de processos químicos. Os indivíduos envolvidos neste trabalho farão também avaliações econômicas de novos processos, projetando equipamentos individuais para a proposição de novos empreendimentos, ou desenvolvendo o esquema de uma planta para coordenação de toda a operação (AERSTIN; STREET, 2012).

Com isso, a habilidade técnica do engenheiro sempre é empregada em conjunto com as estratégias de segurança e o conhecimento econômico.

A existência ou a análise de um processo implica a existência de um produto a ser fabricado e pelos quais certos consumidores estão dispostos a pagar. O produto deve ser entregue numa certa quantidade, satisfazendo certas exigências de qualidade, e a um preço aceitável pelo consumidor. Simultaneamente, o produto deve cobrir o custo dos materiais, da mão de obra e do equipamento, envolvidos na fabricação, e contribuir ainda para um lucro superposto a todos os custos. Para um produto completamente novo, é preciso estabelecer certas estimativas do volume do mercado e a fábrica deve ser dimensionada proporcionalmente a este volume (FOUST, 1981).

A responsabilidade do engenheiro químico ao desenvolver e entregar o processo químico industrial impacta diretamente no sucesso do empreendimento e no retorno do trabalho desempenhado por todos os envolvidos naquele projeto. Engenharia química, quase por definição, envolve a mistura prática e econômica de ciência, engenharia e matemática para resolver problemas e tomar decisões na condução do negócio do empregador (GARRETT, 2012).

Em seu livro Foust (1981), comenta sobre a importância fundamental da matemática como ferramenta do engenheiro ao utilizar os métodos tradicionais de estudo no desenvolvimento de processos, pois é um trabalho quantitativo. Além disso, explica como as atuações dos engenheiros químico e mecânico se misturam nas responsabilidades do trabalho sobre processos e equipamentos, uma vez que utilizando os métodos tradicionais, o engenheiro químico deve ter um conhecimento profundo e extenso sobre a mecânica dos materiais.

4.2 PROCESSOS DE PRODUÇÃO DO BIODIESEL E DO METANOL

Os processos de produção de biocombustíveis tem apresentado importância crescente devido ao esgotamento de reservas de combustíveis fósseis e os impactos ambientais causados pelo uso destas energias não renováveis. Além disso, estima-se que o preço dos combustíveis fósseis sempre aumentará, pois a produção não suprirá a demanda devido ao esgotamento (ASSOCIATION. . . , 2004).

Para Destouni e Frank (2010), até 2050, a porcentagem de fontes de energias renováveis consumidas globalmente vai alcançar até 35% do total de fontes energéticas utilizadas. Esta é uma previsão positiva para o Brasil, que é um dos maiores produtores de biodiesel do mundo, principalmente devido a programas federais de incentivo a essa prática (BIODIESELBR.COM, 2014).

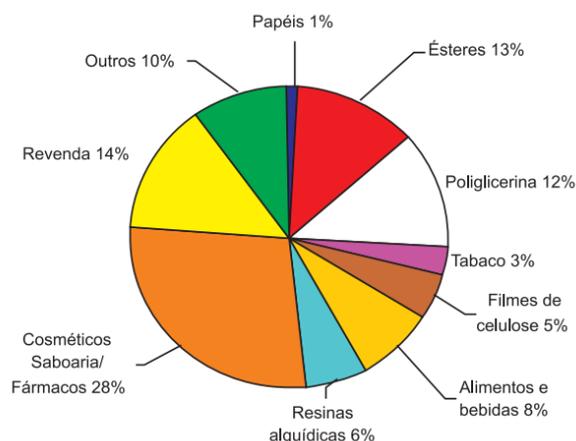
O processo de produção do biodiesel, se dá a partir de biomassa que é capaz de gerar um composto com características energéticas semelhantes ao diesel fóssil comum, que é mais nocivo a saúde e ao meio ambiente quando queimado. Este processo gera como subproduto o glicerol que, em uma das etapas do processo, precisará ser removido (STANKIEWICZ; MOULIJN, 2002). O resíduo de glicerol gerado atinge a faixa das milhões de toneladas por ano e a busca por soluções razoáveis para sua utilização tem se tornado um tópico urgente de um ponto de vista econômico, ambiental e social (QUISPE; CORONADO; JR, 2013).

Dados de produção de glicerol vem apresentando um consumo pelo mercado brasileiro de apenas 10% de todo glicerol produzido no país (VIANA, 2011; UMPIERRE; MACHADO, 2013). Viana (2011) cita os malefícios que o glicerol pode causar sendo ele um material líquido altamente poluente: problemas de intoxicação, formação de espumas em corpos hídricos, mau cheiro e impactos naturais em determinados ecossistemas.

Mota, Silva e Gonçalves (2009) relacionam os principais usos do glicerol depois de purificado e transformado em glicerina, a figura 1 apresenta um gráfico com os principais

setores industriais que consomem a glicerina. O autor enfatiza que em termos de transformação do glicerol, este componente tem aplicações limitadas sendo as principais na produção de explosivos, como a nitroglicerina, e na formação de resinas alquídicas.

Figura 1 – Principais setores industriais de utilização da glicerina



Fonte: Mota, Silva e Gonçalves (2009)

Estudos tem mostrado que o glicerol é uma excelente alternativa de solvente na separação de algumas misturas de álcoois, seguindo assim, os princípios da química verde (ou química sustentável) (KISS; IGNAT, 2012) e (LENARDÃO et al., 2003). Um exemplo do uso do glicerol como solvente é na destilação do metanol produzido a partir de gás natural, como solvente ele apresenta boa compatibilidade com compostos orgânicos, alto ponto de ebulição e baixa agressividade ao meio ambiente (GU; JÉRÔME, 2010).

Recentemente a produção do metanol tem ganhado espaço nos assuntos de pauta ambiental, uma vez que existe um movimento político internacional para classificação do gás natural, que dá origem ao metanol, como energia “verde” (KRUKOWSKA; AINGER, 2022).

No Brasil existe uma disponibilidade crescente de gás natural e segundo estimativas da Empresa Brasileira de Pesquisa Energética (EPE), até 2029, a disponibilidade de gás natural deve mais do que duplicar, atingindo os 138 milhões de metros cúbicos por dia (LEÃO, 2020).

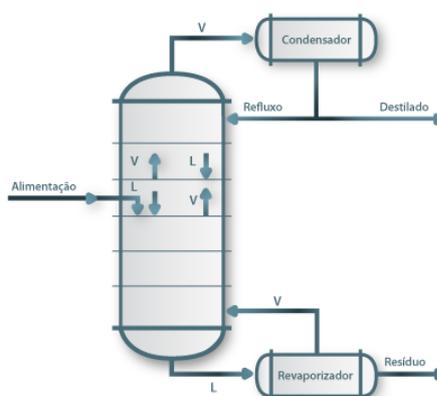
A combinação destas duas situações dentro da industria química nos chama a atenção para o processo que pode solucionar o excesso de subproduto da produção de biodiesel e aproveitar sua disponibilidade na extração do metanol proveniente do gás natural.

A utilização do glicerol excedente como solvente na produção do metanol a partir do gás natural é uma alternativa interessante se considerarmos que a produção do biodiesel também pode fazer uso do metanol como solvente (PEREIRA et al., 2016). Existe uma sinergia entre os dois sistemas de produção onde o produto gerado por um dos processos pode ser usado como solvente no outro processo.

4.3 COLUNA DE DESTILAÇÃO

O equipamento onde ocorre a extração do metanol é a coluna de destilação. A coluna de destilação é o equipamento originalmente projetado para a separação dos componentes de uma mistura através do contato entre a fase líquida e o vapor dessa mistura. Como mostra a figura 2, os níveis que compõem o interior da coluna permitem a vazão de líquido descendente e a de vapor ascendente, proporcionando assim o contato entre as fases e a transferência de massa e energia (HOLLAND, 1982). Conectados a coluna com divisões onde ocorre o contato entre as fases, encontram-se um condensador (ligado ao topo) e um refeedor (na base), que promovem respectivamente, a remoção e o reciclo dos componentes. Além de uma saída de resíduo conectada próxima ao refeedor.

Figura 2 – Representação de uma coluna de destilação



Fonte: Murti e Winkle (1958)

A construção de uma coluna de destilação pode ser uma obra de grandes dimensões, com estruturas que atingem mais de 100 metros de comprimento e diâmetros de até 15 metros. As pressões de operação podem variar de 0,02 atm até $3,4 \times 10^6$ atm (HOLLAND, 1982). De acordo com Humphrey e Siebert (1992) e Engeliem e Skogestad (2004) a destilação é o processo de separação mais amplamente utilizado em indústrias químicas e responsável por aproximadamente 3% de todo o consumo de energia do mundo.

4.4 MODELAGEM DE PROCESSOS INDUSTRIAIS

Segundo Perlingeiro (2005), “a Engenharia Química trata das instalações cuja finalidade é transformar substâncias em outras de interesse industrial”. Essa transformação é normalmente complexa e pode exigir várias etapas. Entre as etapas iniciais de um processo está a modelagem, sendo que o modelo é a descrição matemática daquele processo (EDGAR et al., 2001).

A modelagem é uma das principais áreas de estudo da engenharia química, que se define no reconhecimento do processo e posteriormente na sua otimização (FRANKS, 1972). Phelps

(2016) define modelagem e otimização como a melhoria de um processo existente, sua situação, equipamento ou sistema. Ele comenta que é algo complexo, uma prática inacabável que exige investigação criativa de cada *design*.

Edgar et al. (2001) listam em seu livro os conhecimentos básicos exigidos para desenvolvimento de um modelo matemático tradicional, conforme segue: leis fundamentais da física e química, tais como as leis de conservação de massa, energia e quantidade de movimento, e os conceitos de equilíbrio. Com isso o engenheiro químico responsável por modelar um processo ou equipamento deve conhecer as parametrizações do escoamento de fluidos, da transferência de calor, da transferência de massa, a cinética química envolvida, as interações termodinâmicas e os conceitos de controle da operação.

De fato, a modelagem tradicional de processos químicos é uma tarefa complexa, em grandes sistemas estes modelos podem ter alto custo, principalmente se o sistema tiver um comportamento não-linear também complexo (BIKMUKHAMETOV; JÄSCHKE, 2020). Estes modelos matemáticos são chamados "transparentes", pois as relações de entrada e saída do sistema são explícitas e uma mudança na entrada leva a uma alteração na saída (MATZOPOULOS, 2011). Outros modelos, não baseados nas leis fundamentais da física, mas nos conceitos ciência de dados e da inteligência artificial possuem um desenvolvimento menos complexo, mas em contrapartida, as relações de entrada e saída são menos evidentes.

4.4.1 Projeto de uma Coluna de Destilação

Quando se fala no projeto de uma coluna de destilação um dos métodos considerados tradicionais é o método de McCabe-Thiele (GREEN; PERRY, 2008). Este método é baseado em diversas simplificações como:

- a energia para vaporização dos componentes alimentados é a mesma;
- para cada mol de líquido vaporizado, um mol de vapor é condensado;
- a energia térmica da mistura é desprezível.

Além disso, se trata de um método gráfico, onde as constantes de projeto são obtidas pelos desenhos das linhas de operação em esquemas pré-determinados (MCCABE; SMITH; HARRIOTT, 1993). Ao considerar um conjunto de dados com várias opções de entrada e saída no equipamento a análise pode tomar muito tempo.

A modelagem e otimização do processo de destilação considera o custo gerado pelas condições de temperatura e pressão de operação, volume de produto e concentração final do produto. (SCHNEIDER et al., 2014). Considerando o processo de destilação usado na produção de etanol, Teixeira (2003) comenta que a destilação é a etapa mais onerosa e limitante do processo, atinge até 80% do custo operacional industrial por demandar grande fluxo energético.

4.5 MODELOS DE APRENDIZADO DE MÁQUINA

Schweidtmann et al. (2021) explica que Aprendizado de Máquina (AM) é uma subclasse da Inteligência Artificial (IA). Tem raízes na ciência da computação e na matemática e fornece ao computador a habilidade de aprender a partir dos dados sem uma programação explícita. As primeiras aplicações da Inteligência Artificial na Engenharia Química foram propostas com a criação de sistemas especializados, por exemplo, para propriedades termofísicas e *designs* catalíticos na década de 80, mas não alcançaram descobertas principalmente por causa do custo do processamento computacional da época (BANARES-ALCANTARA; WESTERBERG; RYCHENER, 1985).

Estamos em um momento favorável para o uso de modelos de Aprendizado de Máquina. Assim como a Indústria 4.0, com foco no uso de Inteligência Artificial (IA), “Big Data” e a Internet das Coisas (IoT), a Indústria 5.0 incorpora estes sistemas e acrescenta uma inteligência humana ainda maior (SONDH, 2021). Além disso, devido ao uso abrangente de sistemas de controle distribuídos na indústria, uma grande quantidade de dados foi alcançada nas últimas décadas (GE et al., 2017).

Para a utilização de um modelo de aprendizado de máquina são necessários 7 passos: Coletar dados, preparar os dados, escolher o modelo, treinar o modelo, avaliar o modelo, ajustar os parâmetros e, por fim, fazer as previsões (YUFENG, 2017).

Os modelos de aprendizado de máquina são a estrutura lógica dos algoritmos que fazem previsões, classificações e muitas outras tarefas. Kaur (2019) relaciona os principais usos de algoritmos de aprendizado de máquina comentando que se tratam de uma das inovações modernas da humanidade com uso não só em processos industriais como no dia a dia individual.

Pedamkar (2022) lista os principais usos comuns de algoritmos de aprendizado de máquina na atualidade. Sendo eles:

- Reconhecimento: reconhecimento facial no celular e reconhecimento de voz, como por exemplo assistentes pessoais virtuais;
- Previsões: Monitoramento de tráfego de carros, controle de processos industriais, bolsa de valores, entre outros...
- Segurança: Detecção de comportamentos não usuais em áreas monitoradas por câmeras;
- Atendimento ao público: Ferramentas de busca na internet, análise de dados de demandas do cliente, redes sociais, filtro de e-mails, entre outros...

Para indústria podemos citar algumas revisões como de Narciso e Martins (2020) do uso de aprendizado de máquina em publicações sobre eficiência energética. Os autores relatam um aumento drástico no número de publicações sobre o assunto a partir de 2016 sendo os tipos de

indústrias com mais publicações as petroquímicas com concentração de trabalhos originados na China. Thangaraj e Narayanan (2018) comentam em sua publicação "Indústria 1.0 a 4.0: A evolução das fábricas inteligentes" que, de 2018 a 2023, 50% dos investimentos planejados serão destinados a atualização das indústrias tradicionais ao conceito de indústria 4.0.

E Dalzochio et al. (2020) revisam o uso de modelos de aprendizado de máquina para manutenção preditiva na Indústria 4.0, concluindo que a necessidade de dados em grande escala ainda é um obstáculo. Porém, os autores afirmam que ainda existe muito espaço para desenvolver aplicações e continuam frequentemente surgindo novas.

Em contraste aos modelos com base nos princípios fundamentais, os modelos de aprendizado de máquina estimam variáveis diretamente dos dados, explorando a habilidade de encontrar padrões complexos sem prover uma forma explícita para isso. Isto torna os modelos de aprendizado de máquina mais fáceis de serem construídos em comparação aos modelos de princípios fundamentais. No entanto, estes modelos exigem uma grande quantidade de dados para encontrar os padrões de forma precisa e, por não proverem uma forma explícita para isso, possuem uma natureza chamada de "caixa-preta" que pode dificultar o diagnóstico de erros no modelo (ROSCHE et al., 2020).

Algoritmos de aprendizado de máquina podem ser classificados em três categorias: aprendizado supervisionado, aprendizado não-supervisionado e aprendizado de reforço. Nos casos em que uma propriedade (rótulo¹) está disponível para o conjunto de dados de treinamento, mas está faltando para outras instâncias² é útil utilizar algoritmos de aprendizado supervisionado. Agora, se é preciso descobrir as relações implícitas em um conjunto de dados não marcados é mais interessante utilizar algoritmos de aprendizado não-supervisionado. Quando há algum tipo de *feedback* disponível para cada passo ou ação preditiva, mas sem rótulo preciso ou mensagem de erro é quando utiliza-se algoritmos de aprendizado de reforço (LE, 2016).

Para modelagem de equipamentos industriais, como a coluna de destilação, onde os dados trazem rótulos como a fração do destilado no produto utilizam-se algoritmos supervisionados.

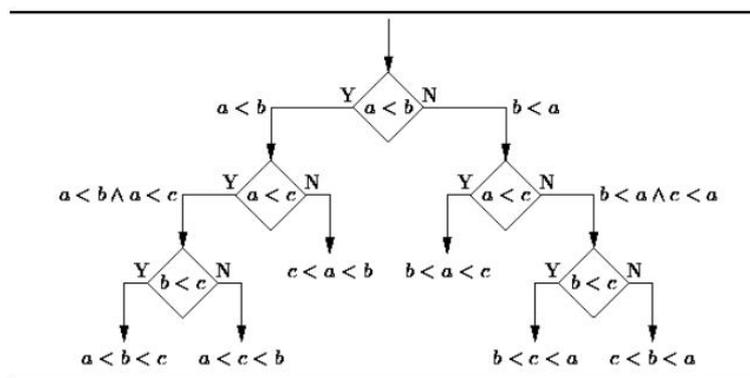
Um exemplo de algoritmo de aprendizado supervisionado é o algoritmo de Árvores de Decisão, que de maneira simplificada é um número mínimo de perguntas sim/não que são precisas para estimar a probabilidade de se tomar uma decisão correta com maior frequência. Com esta probabilidade pode-se atacar o problema de forma estruturada e sistemática chegando a uma conclusão lógica (LE, 2016). Quando se utiliza um banco de dados de diversas variações de um sistema de produção industrial, este algoritmo pode ser uma boa opção para seleção das opções viáveis. A figura 3 apresenta uma representação de como funciona essa sequência de perguntas e decisões.

Outro algoritmo supervisionado é o chamado Floresta Aleatória ou Floresta Aleatória de

¹ rótulo é a variável de saída, que o modelo tenta prever

² instâncias são as assertões dos valores dos dados no modelo, por exemplo: treinamento, teste

Figura 3 – Representação do funcionamento do algoritmo de árvores de decisão

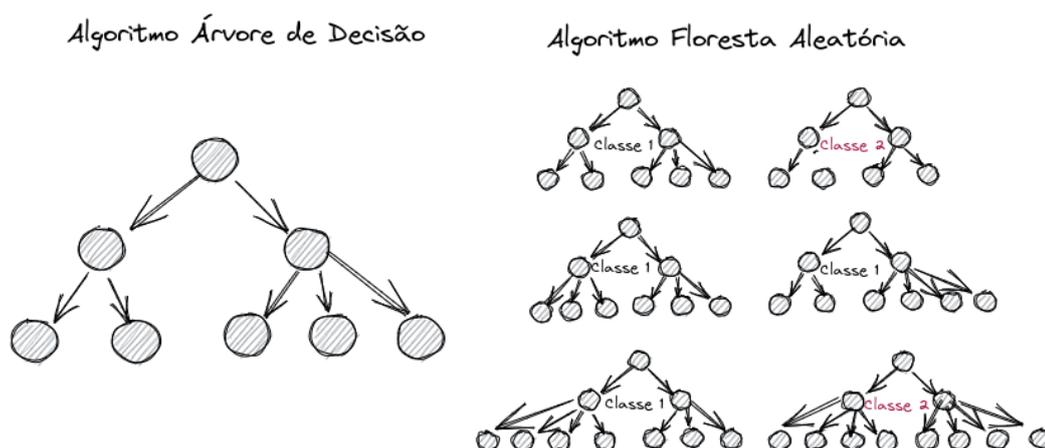


Fonte: Le (2016).

Decisão que é formado por um conjunto de algoritmos de árvore de decisão. É um método que pode ser utilizado para classificação ou regressão e tem a característica de entregar os dados de forma aleatória antes da etapa de treinamento do algoritmo. No caso de classificação, a saída do algoritmo é a classe mais escolhida pela maioria das árvores de decisão. No caso da regressão, a resposta é a média dos valores previstos por cada árvore (HO, 1995).

A figura 4 demonstra como um algoritmo compõe o outro, e como é feita a decisão do algoritmo floresta aleatória em uma aplicação de classificação. A classe mais "votada" pelos algoritmos de árvore de decisão que compõem o floresta aleatória é a saída do algoritmo.

Figura 4 – Comparação entre árvore de decisão e floresta de decisão



Fonte: Elaborado pelo autor, 2022

O algoritmo de floresta aleatória em geral apresenta maior precisão e é melhor porque corrige o hábito de "overfitting" que o árvore de decisão costuma apresentar. Porém, as características dos dados podem afetar sua performance (HASTIE et al., 2009).

Para conjuntos de dados com fortes relações lineares entre as variáveis de controle e as variáveis resposta, alguns processos podem ser modelados por algoritmos mais simples. Para

verificar essa linearidade pode se aplicar o ajuste ao modelo de regressão linear. Regressão linear é um algoritmo de aprendizado supervisionado que executa a tarefa de prever o valor da variável alvo (y) baseado no valor dado da variável de controle (x), buscando os coeficientes da linha de regressão que mais aproxima o valor predito do valor real (MAAOUANE et al., 2021).

A equação a seguir exemplifica uma função hipotética que compõe o algoritmo de regressão linear.

$$y = a + b.x$$

Dado valor de " x " o algoritmo busca os coeficientes " a " e " b " de menor desvio da raiz quadrada do erro médio (RMSE), de forma a aproximar o valor " y " da saída esperada.

4.6 CIÊNCIA DE DADOS NA ENGENHARIA QUÍMICA

MacMurray e Himmelblau (1995) relatam em seu livro que o desenvolvimento de modelos de projeto através da ciência de dados é uma prática da engenharia química desde a década de 80, com a publicação especialmente na década de 90 dos principais trabalhos que utilizavam algoritmos simples e obtinham resultados promissores.

Trabalhos ainda mais promissores de estimativa da composição de produtos de colunas de destilação apareceram entre os anos 2000 e 2010, onde modelos de redes neurais eram alimentados com dados de simulação computacional (SINGH; GUPTA; GUPTA, 2005). E depois ainda, trabalhos baseados no algoritmo de Levenberg-Marquardt trouxeram resultados ainda mais eficientes para treinar modelos de otimização (SINGH; GUPTA; GUPTA, 2007).

Outras publicações, de conteúdo semelhante aos trabalhos de Singh et al. 2005 e 2007, utilizam dados coletados de uma coluna experimental e os denominam termo sensores virtuais (CANETE et al., 2012).

Kiss (2011) comparou dois estudos de simulação de colunas de destilação de sua própria autoria com auxílio do software Aspen Plus e técnicas de ciência de dados e inteligência artificial e obteve resultados positivos: na simulação em que utilizou a estratégia de sistema de calor integrado, onde as correntes de entrada e saída realizam a troca térmica, em comparação a uma simulação onde não aplicava este sistema, e alcançou uma redução de 47% no gasto com resfriamento e 43% no gasto com aquecimento.

Lourenço (2021) em uma revisão recente de trabalhos publicados nas grandes subáreas da engenharia química: cinética e reatores, controle de processos, termodinâmica, fenômenos de transporte e operações unitárias aplicando ferramentas da ciência de dados, analisou 307 documentos no período entre 1992 e 2021, e constatou uma taxa de crescimento anual das publicações de 22,25% principalmente em países como EUA e China. No âmbito nacional

identificou poucos trabalhos publicados na área em comparação com a expansão internacional, o que evidencia oportunidades de desenvolvimento de projetos de pesquisa inovadores.

5 MATERIAIS E MÉTODOS

5.1 FERRAMENTAS UTILIZADAS

5.1.1 Ferramentas para Obtenção dos Dados

A fonte de dados deste trabalho foi uma coluna de destilação extrativa simulada em computador. Para realizar esta simulação utilizaram-se os módulos desenvolvidos por Cortes-Peña et al. (2020) no simulador de código aberto "BioSTEAM", composto por diversos módulos gratuitos e públicos de simulação de biorrefinarias e análises tecno-econômicas. O anexo A apresenta a página da documentação deste simulador e seu repositório aberto no GitHub.

Para inserção dos dados de entrada no simulador, captura dos dados de saída e registro de todos os valores em arquivo de trabalho foram utilizadas: a linguagem de programação Python, as bibliotecas Python "secrets", "choice", "numpy" e "csv". Além do *software* Visual Studio Code para edição de código.

Os endereços da documentação destas ferramentas são apresentados no anexo B.

5.1.2 Ferramentas de Modelagem por Aprendizado de Máquina

Para modelagem por aprendizado de máquina os dados gerados na simulação tiveram o ajuste avaliado nos modelos de aprendizado supervisionado citados por Le (2016), Ho (1995), Hastie et al. (2009) e (MAAOUANE et al., 2021), foram eles:

- Modelo de Regressão Linear;
- Modelo de Floresta Aleatória;

O software usado para avaliação dos modelos foi o "Orange Data Mining" cuja interface gráfica e endereço da documentação são apresentados no anexo C.

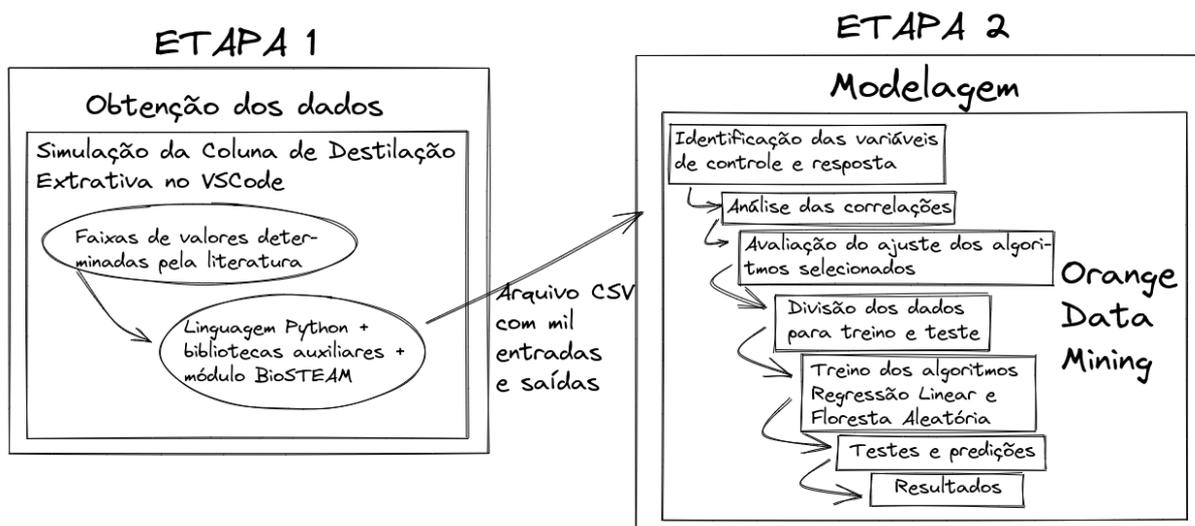
5.2 MÉTODO DE MODELAGEM POR APRENDIZADO DE MÁQUINA

5.2.1 Diagrama do Método Utilizado

O método utilizado foi dividido em duas grandes etapas, obtenção dos dados e modelagem, a figura 5 ilustra a divisão de tarefas para obtenção dos resultados deste trabalho, cada bloco contém suas etapas envolvidas pelas ferramentas utilizadas.

Cada uma das etapas ilustradas serão detalhadas nos tópicos a seguir. Após seleção do processo, pesquisa literária e escolha das ferramentas o método se torna efetivamente rápido,

Figura 5 – Diagrama de divisão de tarefas do método



Fonte: Elaborado pelo autor, 2022

uma vez que depende apenas de processamento computacional, e a capacidade de processamento de dados dos processadores atuais é de milhares de megabytes por segundo (ALMEIDA, 2016).

5.2.2 Etapa de Simulação

Para realizar a simulação foi necessário entender os limites do módulo BioSTEAM de simulação da coluna, e também as faixas de funcionamento de colunas de destilação extrativas usadas em outros trabalhos. O quadro 1 apresenta os valores determinados para simulação e algumas literaturas usadas como base.

Quadro 1 – Faixas de operação da coluna de destilação extrativa simulada

Variável	Faixa	Unidade	Referência
RR	1.1 - 7	-	Cortes-Peña et al. (2020) e Errico e Rong (2012)
Vazão de Água	3,42 - 17,1	kmol/h	Franchi (2010) e Kiss e Ignat (2012)
Vazão de Metanol	30 - 150	kmol/h	Cortes-Peña et al. (2020) e Kiss e Ignat (2012)
Vazão de Glicerol	30 - 150	kmol/h	Dong, Dai e Lei (2018)
Temperatura	40 - 88	°C	Douglas e Hoadley (2006)
Pressão	101,325 - 200	kPa	Cortes-Peña et al. (2020) e Kiss e Ignat (2012)

Fonte: Elaborado pelo autor, 2022

Analisando as vazões de alimentação e solvente no quadro 1 nota-se as faixas de valores idênticas para metanol e glicerol, e quantidades baixas para água. Esse é o padrão identificado na literatura Dunford (2007) e Ignat e Kiss (2012), onde a quantidade de solvente indicada é sempre similar ou superior a quantidade do álcool. Para a simulação, cada valor aleatório determinado para metanol recebe como entrada um valor idêntico de solvente.

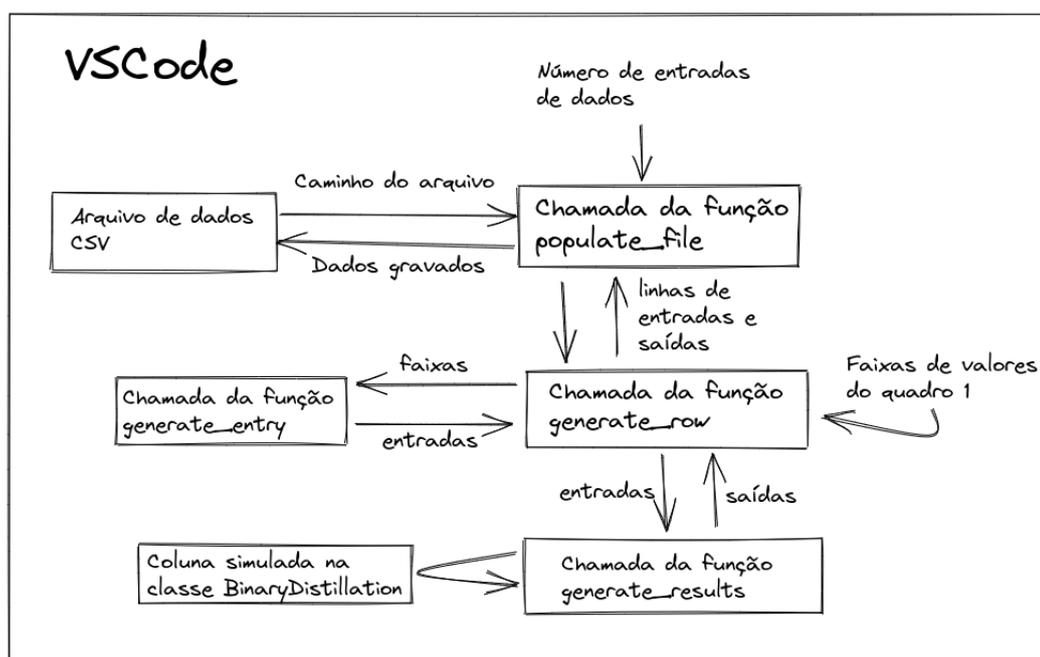
A faixa de valores determinada para razão de refluxo (RR) teve seu limite inferior limitado pela capacidade do simulador (CORTES-PEÑA et al., 2020) e o limite superior com base nos trabalhos de Errico e Rong (2012), Dong, Dai e Lei (2018), Douglas e Hoadley (2006) e Zhang, Liang e Feng (2010). Estes trabalhos também serviram de referências de faixas de valores de temperatura e pressão de operação.

As variáveis apresentadas no quadro 1 são também as variáveis de controle do modelo de aprendizado de máquina na etapa seguinte, e a variável resposta será o custo total de instalação do equipamento, uma vez que o módulo BioSTEAM tem sua avaliação tecno-econômica reconhecida na literatura (SHI; GUEST, 2020; SANCHIS-SEBASTIÁ et al., 2020).

Deste modo, buscou-se uma modelagem específica da coluna de destilação com ênfase em suas variáveis de projeto e em condições de resposta onde a relação entre a entrada e a saída fosse possivelmente pouco linear. Os resultados da modelagem mostrarão a seguir essas relações.

As etapas de simulação do código escrito e apresentado nos apêndices A, B e C são representadas na figura 6.

Figura 6 – Etapas realizadas pelo código de simulação



Fonte: Elaborado pelo autor, 2022

O código mostrado no apêndice A apresenta a função de geração de resultados da coluna de destilação extrativa que se fundamenta na classe “BinaryDistillation” do módulo “Distillation” da biblioteca BioSTEAM. Esta função é abastecida pelos valores da função de geração de linha de dados, intitulada “generate_row” no apêndice B, que é suprida pela função de geração de valores de entrada chamada “generate_entry” e pelas faixas de valores do quadro 1.

Todos os valores de entradas e saídas são gravados em um arquivo “.csv” para utilização

no software "Orange Data Mining" na etapa de modelagem da coluna de destilação.

As funções foram escritas de modo que fosse possível alterar:

- As saídas da simulação da coluna de destilação;
- As faixas de valores de entrada na coluna;
- O caminho para outros arquivos de dados;
- E o número de linhas de entradas e saídas gravadas.

Além disso, cada função tem uma responsabilidade única para facilitar a manutenção ou atualização do código.

5.2.3 Etapa de Modelagem por Aprendizado de Máquina

Todas as etapas de modelagem aconteceram dentro da plataforma interativa do software "Orange Data Mining" apresentada no anexo C.

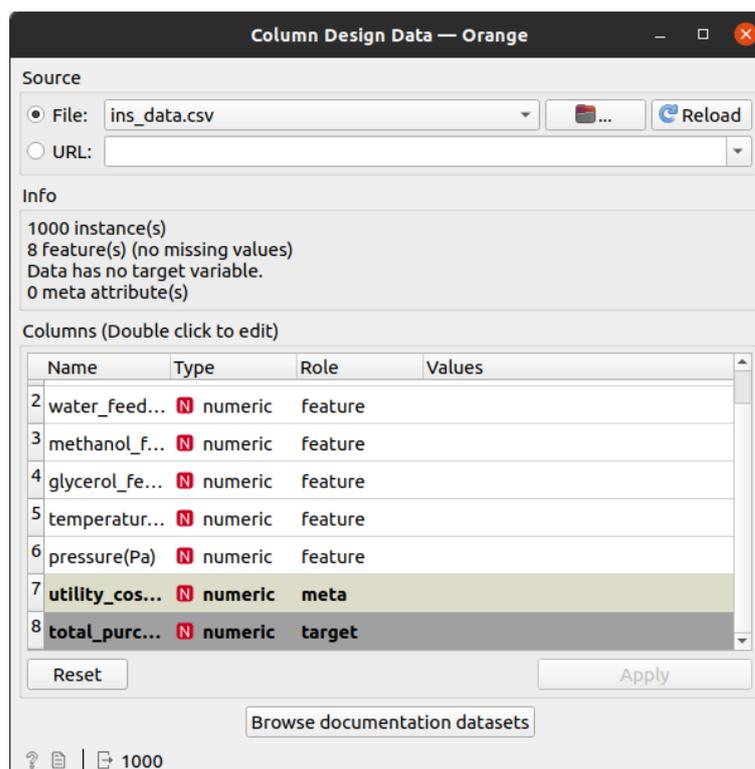
A sequência de trabalho da modelagem da coluna de destilação extrativa foi dividida nas seguintes etapas:

1. Importação do arquivo de dados;
2. Verificação do arquivo importado em tabela;
3. Análise da correlação linear entre as variáveis;
4. Análise estatística das propriedades;
5. Teste de ajuste de modelos e comparação entre eles;
6. Separação dos dados em conjuntos de treino e de teste;
7. Treinamento dos modelos;
8. Teste de precisão das predições;
9. Análise por gráfico de dispersão;
10. Visualização dos coeficientes gerados pelo modelo;
11. Armazenamento dos dados e do modelo;

Um esquema de trabalho foi desenvolvido dentro do software “Orange Data Mining” com as conexões de cada etapa na sequência de operação determinada pelo programa.

O primeiro passo foi a importação do arquivo de dados e nele selecionadas as variáveis de funcionalidade e a variável alvo como mostra a figura 7. Foram verificados as entradas e rótulos importados e em seguida feitas as análises de correlações lineares pelo método de Pearson.

Figura 7 – Identificação das variáveis de funcionalidade e alvo no *Orange*.



Fonte: Elaborado pelo autor, 2022

Extraíram-se os valores estatísticos dos dados simulados como média, mediana, dispersão, mínimos e máximos. Então, cada modelo de aprendizado de máquina foi submetido aos dados da simulação para verificação do ajuste pelas métricas MSE¹, RMSE², MAE³ e R²⁴.

Os dados foram divididos em 70% para treinamento dos algoritmos e o resto para teste. A proporção de dados dividida para treino e teste é estabelecida pela literatura e muito semelhante para todos os casos de uso de modelos de aprendizado de máquina (YUFENG, 2017). Durante o treinamento selecionou-se o número de 10 árvores de decisão para o algoritmo floresta aleatória, esse número foi escolhido porque valores superiores a este não exerciam influência nos valores das métricas de ajuste da etapa anterior.

¹ Erro Médio Quadrático

² Raíz Quadrada do Erro Médio

³ Erro Absoluto Médio

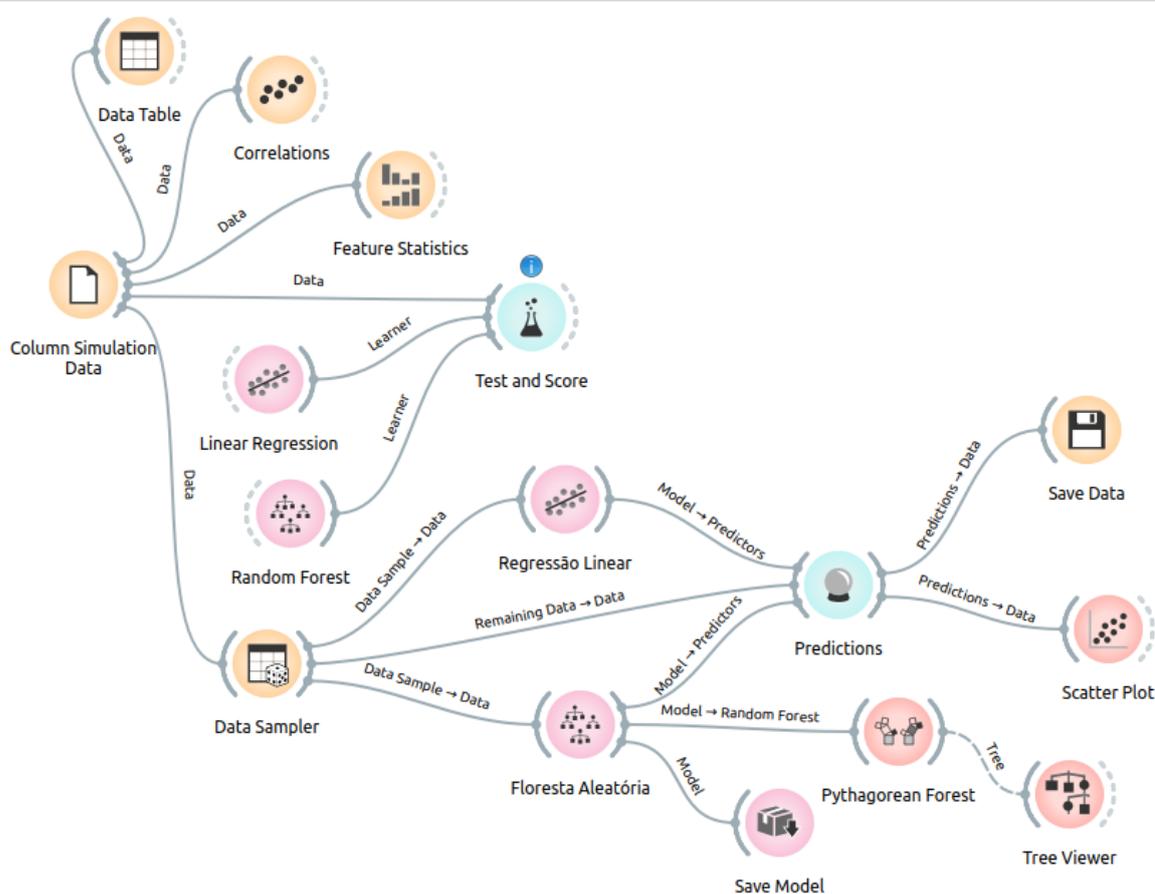
⁴ Coeficiente de Determinação

Por fim, foram realizados os testes e os valores previstos pelos modelos foram analisados em tabela e gráfico de dispersão apresentados na seção de resultados.

Para auxiliar na análise da adaptação do modelo floresta aleatória, foram geradas visualizações das árvores de decisão formadas.

A figura 8 apresenta o esquema completo da modelagem como ela é apresentada no *software*. Os resultados gerados foram exportados em um arquivo do “Orange Data Mining” para o modelo completo, e em arquivo “.xlsx” dos dados para acesso em ambiente de “Microsoft Office” ou “Libre Office”.

Figura 8 – Esquema de trabalho completo no *Orange*.



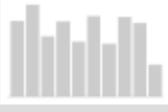
Fonte: Elaborado pelo autor, 2022

6 RESULTADOS E DISCUSSÃO

6.1 ANÁLISE EXPLORATÓRIA DOS DADOS

A figura 9 apresenta os valores estatísticos dos dados simulados e uma prévia da distribuição. O custo total é expresso em dólares americanos (US\$).

Figura 9 – Valores estatísticos dos dados simulados.

Name	Distribution	Mean	Median	Dispersion	Min.	Max.
Vazão de água (kmol/h)		10.163	10	0.395161	3	17
Razão de refluxo		3.9431	3.95	0.444707	1.1	6.9
Vazão de metanol (kmol/h)		88.915	90	0.393554	30	145
Vazão de glicerol (kmol/h)		88.915	90	0.393554	30	145
Temperatura (°C)		62.986	62	0.221513	40	86
Pressão (atm)		151443	151325	0.189974	101325	199325
Custo total (US\$)		189928	185745	0.129677	139698	263058

Fonte: Elaborado pelo autor, 2022

O apêndice D contém os histogramas dos 200 primeiros valores de cada propriedade simulada.

Todas as propriedades apresentaram valores nos extremos determinados pela simulação e com boa dispersão ao longo das faixas definidas, o que indica que não há presença de *outliers*¹ no banco de dados.

¹ Em estatística, outlier, valor aberrante ou valor atípico, é uma observação que apresenta um grande afastamento das demais da série, ou que é inconsistente.

A figura 10 apresenta os valores de correlação linear de Pearson para cada propriedade em comparação com o custo total da coluna de destilação.

Figura 10 – Valores da correlação linear de Pearson.

Pearson correlation

1	+0.815	Custo total	Vazão de glicerol
2	+0.815	Custo total	Vazão de metanol
3	+0.815	Custo total	Vazão de água
4	+0.458	Custo total	Razão de refluxo
5	-0.016	Custo total	Pressão
6	+0.004	Custo total	Temperatura

Fonte: Elaborado pelo autor, 2022

Com exceção da pressão e da temperatura, todas as propriedades apresentaram um coeficiente de correlação linear de moderada a forte diretamente proporcional com o custo total da coluna. Como foram utilizadas medidas de pressão e temperatura dentro da faixa de operação normal da coluna, espera-se que essas propriedades não realizem muita influencia sobre o custo total da operação, uma vez que apenas em pressões e temperaturas extremas o custo de instalação da coluna teria variação significativa pela necessidade de seleção de material resistente.

Os coeficientes de correlação com valores altos e semelhantes apresentados nas vazões estão coerentes, pois o aumento ou diminuição das vazões interferem diretamente no volume da coluna e conseqüentemente no seu custo, com pequenas ressalvas em faixas próximas que um volume de colunas poderia suportar tendo as mesmas dimensões.

A razão de refluxo apresentou efeito moderado, o que também era esperado pois o valor é calculado sobre o refluxo mínimo que a coluna deve possuir. Dessa forma, somente variações muito altas desse valor refletiriam nas dimensões do equipamento e, dessa forma, no custo total de instalação.

6.2 AJUSTE DOS MODELOS SELECIONADOS

Como citado anteriormente, foram selecionados os modelos de aprendizado de máquina Regressão Linear e Floresta Aleatória. Antes do treinamento e aplicação dos modelos efetivamente, testes preliminares de ajuste aos dados foram realizados. O modelo de regressão linear foi selecionado para demonstrar o grau de linearidade dos dados e, se não-linear, justificaria a utilização do segundo modelo.

Para avaliação do grau de ajuste foram escolhidos quatro métricas estatísticas: MSE, RMSE, MAE e R^2 .

O quadro 11 apresenta os valores de cada métrica em comparação para os dois modelos.

Figura 11 – Valores de comparação do ajuste preliminar de cada modelo de aprendizado de máquina.

Model	MSE	RMSE	MAE	R2
Regressão Linear	90393103.323	9507.529	7452.393	0.851
Floresta Aleatória	11048675.436	3323.955	2441.857	0.982

Fonte: Elaborado pelo autor, 2022

Pode-se perceber que em todas as métricas o modelo floresta aleatória obteve melhor ajuste, mostrando que a interação não-linear entre as entradas e as saídas é mais forte que a interação linear, validando a necessidade de utilização do modelo Floresta Aleatória. Esta característica reforça o efeito não linear das propriedades temperatura, pressão e razão de refluxo apresentadas na figura 10 com valores de coeficiente de correlação linear de Pearson inferiores.

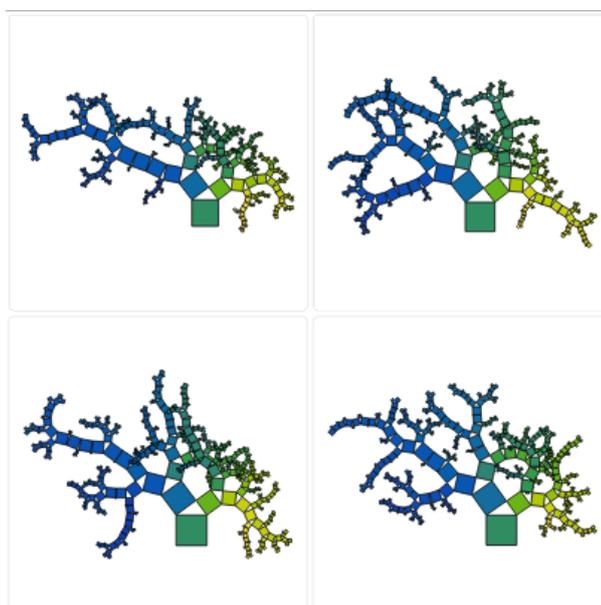
Podemos notar que as métricas MSE e RMSE obtiveram valores muito maiores que a MAE em ambos os modelos, logo podemos concluir que existem valores relevantes de grande desvio da média. Por esse motivo também o ajuste do modelo floresta aleatória teve melhor desempenho que o modelo regressão linear, valores dispersos e que influenciam o modelo impossibilitam uma representação por reta.

6.3 ÁRVORES DE DECISÃO GERADAS

As dez árvores de decisão geradas pelo algoritmo Floresta de Decisão podem fornecer algumas informações sobre quais faixas de valores os dados apresentam menor comportamento linear para comparação com os valores previstos.

Observando a figura 12 onde a intensidade de cor varia com a média dos valores nos ramos, pode-se identificar uma tendência de extensão dos ramos em direção as regiões azuladas. Selecionando uma dessas árvores, aproximando a visualização nos seus primeiros nós e analisando para quais faixas de valores essas áreas migram, pode-se ver que é em relação a modelagem de colunas com menor vazão e volume. Isso demonstra a adaptação do modelo para a característica fortemente não linear dos dados nessas áreas, gerando um número maior de ramificações.

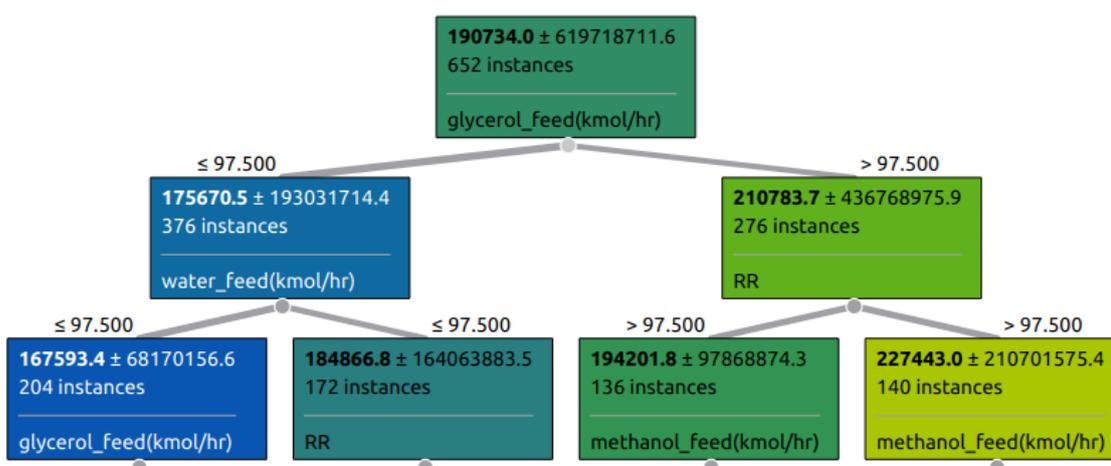
Figura 12 – Tendência das médias nas árvores de decisões geradas.



Fonte: Elaborado pelo autor, 2022

A representação individual da árvore de decisão apresenta os coeficientes do modelo para aquela árvore, a quantidade de instâncias que foram filtradas, a propriedade avaliada e o valor dessa propriedade. Como mostra a figura 13.

Figura 13 – Aproximação nos primeiros nós



Fonte: Elaborado pelo autor, 2022

A seguir verificaremos os resultados previstos pelos modelos e compararemos o gráfico de dispersão destas áreas de modelagem de colunas com menor vazão e volume.

6.4 PREDIÇÕES DO MODELO

Foram realizadas previsões de ambos os modelos para 385 valores selecionados para teste dos modelos. O quadro 14 apresenta a seleção de dez valores de custo total de instalação em dólares americanos (U\$) da coluna de destilação de cada modelo entre mínimos, intermediários e máximos igualmente espaçados entre a lista dos 385 totais.

Figura 14 – Valores de previsão do resultado de cada modelo de aprendizado de máquina.

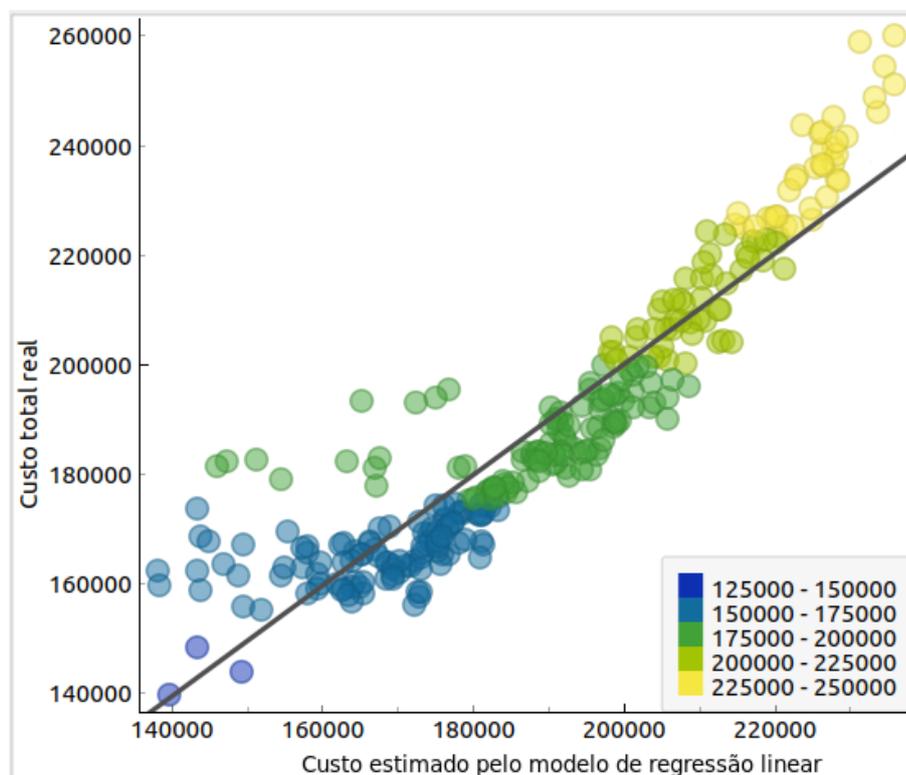
	Custo total real	Floresta aleatória	Regressão linear
1	177518	178167	182743
2	173270	169719	178147
3	161021	159581	169307
4	163779	164539	146828
5	165546	163581	176780
6	184316	185503	191435
7	177518	178167	182743
8	174220	174344	175950
9	233730	235261	228302
10	192176	192531	201178

Fonte: Elaborado pelo autor, 2022

Pode-se observar que em alguns valores a previsão pelo modelo de regressão linear variou em até 8% em relação ao valor real. O desvio padrão dos valores reais foi de U\$ 19.940,14, dos valores previstos pelo modelo floresta aleatória o desvio foi de U\$ 20.762,52 e do modelo de regressão linear foi de U\$ 21.151,20.

A figura 15 apresenta a distribuição gráfica dos valores previstos pelo modelo de regressão linear em relação aos valores reais.

Figura 15 – Gráfico de dispersão do modelo de regressão linear



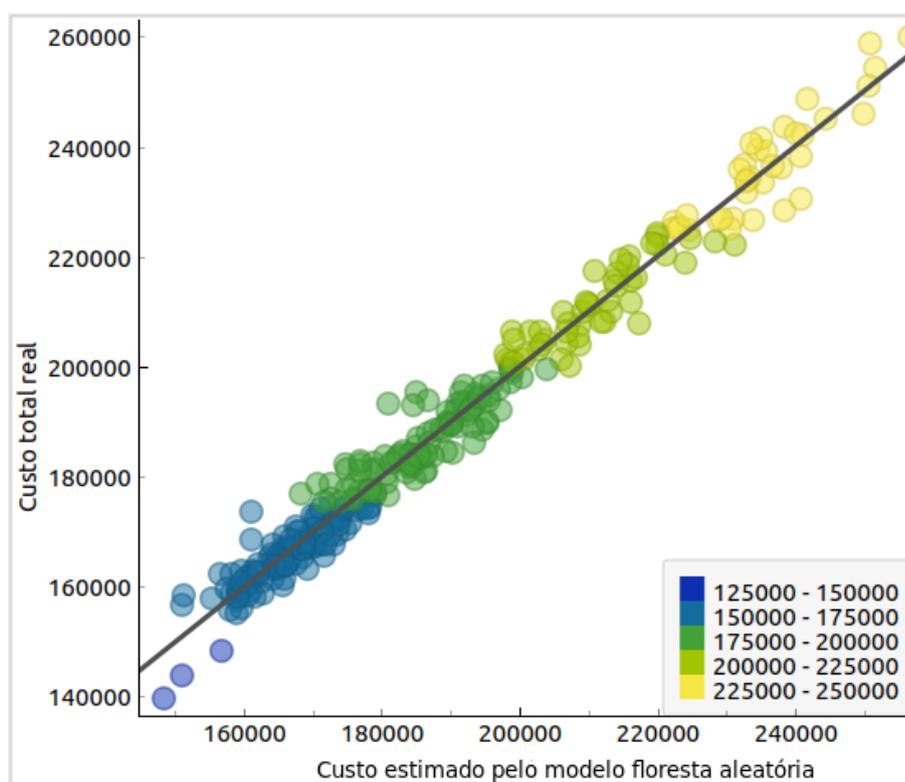
Fonte: Elaborado pelo autor, 2022

Observa-se uma distribuição menos uniforme nas faixas de valores mais baixos, isso pode se dar pelo fato de que os custos de instalação podem apresentar valores fixos mínimos mesmo que as colunas sejam para vazões menores uma vez que há a necessidade de se moldar, toronar e fabricar o material da mesma forma. Estas condições trazem situações de interação não-linear em relação ao custos e as entradas e por esse motivo o modelo de regressão linear não consegue se ajustar com precisão.

Conforme observamos anteriormente nas árvores de decisão geradas pelo modelo de floresta aleatória, notamos um avanço nos ramos (ou nós) em direção a valores médios inferiores. Isso pode representar uma evolução do algoritmo pra tratar essa não linearidade dos dados nessas faixas.

A figura 16 apresenta a distribuição dos valores previstos pelo modelo de floresta aleatória em relação aos valores reais.

Figura 16 – Gráfico de dispersão do modelo de floresta aleatória



Fonte: Elaborado pelo autor, 2022

Pode-se observar que o algoritmo se ajustou as faixas de valores com menor interação linear conseguindo respostas mais próximas ao valor real.

7 CONSIDERAÇÕES FINAIS

A ciência de dados na modelagem de processos químicos industriais ainda é uma estratégia nova e não se conhece o efeito prático da aplicação de todas as suas ferramentas.

Neste trabalho buscou-se aplicar alguns modelos exclusivos da área de aprendizado de máquina a um determinado equipamento e processo.

Nota-se, pelo ajuste dos modelos, a dependência destes a quantidade fornecida de dados, assim como dita a literatura. Além disso, existem muitos outros modelos que podem ajustar-se ainda melhor a este processo e valem a investigação.

Confirma-se que a complexidade de processos químicos industriais exige a aplicação de modelos adaptáveis e não-lineares. O modelo de aprendizado de máquina Floresta Aleatória apresentou melhor ajuste e previsões que o modelo Regressão Linear.

Notavelmente a complexidade e demanda de tempo para o uso destes modelos, uma vez que se conheçam as condições do processo e o modelo que se deseja aplicar, é muito menor que para os modelos tradicionais.

Recomenda-se o usos destes modelos como entradas para modelos de otimização, para tomada de decisão e previsões rápidas acerca de determinado processo.

REFERÊNCIAS

- AERSTIN, F.; STREET, G. *Applied Chemical Process Design*. [S.l.]: Springer Science & Business Media, 2012. Citado na página 16.
- ALMEIDA, R. B. Evolução dos processadores. *Evolução dos processadores*, 2016. Citado na página 27.
- ASSOCIATION of Peak Oil and Gas. 2004. Disponível em: <<http://www.peakoil.net/uhdsg/Default.html>>. Citado na página 17.
- BANARES-ALCANTARA, R.; WESTERBERG, A. W.; RYCHENER, M. D. Development of an expert system for physical property predictions. *Computers & chemical engineering*, Elsevier, v. 9, n. 2, p. 127–142, 1985. Citado na página 21.
- BIKMUKHAMETOV, T.; JÄSCHKE, J. Combining machine learning and process engineering physics towards enhanced accuracy and explainability of data-driven models. *Computers & Chemical Engineering*, Elsevier, v. 138, p. 106834, 2020. Citado 2 vezes nas páginas 13 e 20.
- BIODIESELBR.COM. *Biodiesel no Mundo*. 2014. Disponível em: <<https://www.biodieselbr.com/biodiesel/mundo/biodiesel-no-mundo>>. Citado na página 17.
- CANETE, J. F. de et al. Dual composition control and soft estimation for a pilot distillation column using a neurogenetic design. *Computers & chemical engineering*, Elsevier, v. 40, p. 157–170, 2012. Citado na página 24.
- CORTES-PEÑA, Y. et al. Biosteam: a fast and flexible platform for the design, simulation, and techno-economic analysis of biorefineries under uncertainty. *ACS Sustainable Chemistry & Engineering*, ACS Publications, v. 8, n. 8, p. 3302–3310, 2020. Citado 3 vezes nas páginas 26, 27 e 28.
- DALZOCHIO, J. et al. Machine learning and reasoning for predictive maintenance in industry 4.0: Current status and challenges. *Computers in Industry*, Elsevier, v. 123, p. 103298, 2020. Citado na página 22.
- DESTOUNI, G.; FRANK, H. Source: *Ambio: A journal of the human environment*, 39 (sp1): 18-21. *AMBIO*, BioOne, v. 39, p. 18–21, 2010. Citado na página 17.
- DOGAN, A.; BIRANT, D. Machine learning and data mining in manufacturing. *Expert Systems with Applications*, Elsevier, v. 166, p. 114060, 2021. Citado na página 14.
- DONG, Y.; DAI, C.; LEI, Z. Extractive distillation of methylal/methanol mixture using ethylene glycol as entrainer. *Fluid Phase Equilibria*, Elsevier, v. 462, p. 172–180, 2018. Citado 2 vezes nas páginas 27 e 28.
- DOUGLAS, A. P.; HOADLEY, A. F. A. A process integration approach to the design of the two-and three-column methanol distillation schemes. *Applied thermal engineering*, Elsevier, v. 26, n. 4, p. 338–349, 2006. Citado 2 vezes nas páginas 27 e 28.
- DUNFORD, N. T. *Biodiesel production techniques*. [S.l.]: Oklahoma Cooperative Extension Service, Division of Agricultural Sciences . . . , 2007. Citado na página 27.

- EDGAR, T. F. et al. *Optimization of chemical processes*. [S.l.: s.n.], 2001. Citado 2 vezes nas páginas 19 e 20.
- ENGELIEN, H. K.; SKOGESTAD, S. Selecting appropriate control variables for a heat-integrated distillation system with prefractionator. *Computers & chemical engineering*, Elsevier, v. 28, n. 5, p. 683–691, 2004. Citado na página 19.
- ERRICO, M.; RONG, B.-G. Synthesis of new separation processes for bioethanol production by extractive distillation. *Separation and purification technology*, Elsevier, v. 96, p. 58–67, 2012. Citado 2 vezes nas páginas 27 e 28.
- FOUST, A. S. *Principles of Unit Operations*, (Foust, Alan S.; Wenzel Leonard A.; Clump, Curtis W.; Maus, Lois; Anderson, L. Bryce). [S.l.]: ACS Publications, 1981. Citado 2 vezes nas páginas 16 e 17.
- FRANCHI, C. M. Aplicação de estratégias de controle em coluna de destilação. Universidade Estadual de Maringá, 2010. Citado na página 27.
- FRANKS, R. G. *Modeling and simulation in chemical engineering*. [S.l.: s.n.], 1972. Citado na página 19.
- GARRETT, D. E. *Chemical engineering economics*. [S.l.]: Springer Science & Business Media, 2012. Citado na página 17.
- GE, Z. et al. Data mining and analytics in the process industry: The role of machine learning. *Ieee Access*, IEEE, v. 5, p. 20590–20616, 2017. Citado na página 21.
- GREEN, D. W.; PERRY, R. H. *Perry's chemical engineers' handbook*. [S.l.]: McGraw-Hill Education, 2008. Citado na página 20.
- GU, Y.; JÉRÔME, F. Glycerol as a sustainable solvent for green chemistry. *Green Chemistry*, Royal Society of Chemistry, v. 12, n. 7, p. 1127–1138, 2010. Citado 2 vezes nas páginas 14 e 18.
- HASTIE, T. et al. *The elements of statistical learning: data mining, inference, and prediction*. [S.l.]: Springer, 2009. v. 2. Citado 2 vezes nas páginas 23 e 26.
- HO, T. K. Random decision forests. In: IEEE. *Proceedings of 3rd international conference on document analysis and recognition*. [S.l.], 1995. v. 1, p. 278–282. Citado 2 vezes nas páginas 23 e 26.
- HOLLAND, C. D. Developments in the analysis of multicomponent distillation problems. *Reviews in Chemical Engineering*, De Gruyter, v. 1, n. 2, p. 121–178, 1982. Citado na página 19.
- HUMPHREY, J. L.; SIEBERT, A. F. Separation technologies; an opportunity for energy savings. *Chemical Engineering Progress;(United States)*, v. 88, n. 3, 1992. Citado na página 19.
- IGNAT, R. M.; KISS, A. A. Energy efficient recovery of methanol and glycerol in biodiesel production. *CHEMICAL ENGINEERING*, v. 29, 2012. Citado na página 27.
- INSPECTIONEERING. *Overview of Chemical Processing Industry*. 2021. Disponível em: <<https://inspectioneering.com/tag/chemicals>>. Citado na página 16.

- KAUR, M. *Top 10 real-life examples of Machine Learning*. 2019. Disponível em: <<https://bigdata-madesimple.com/top-10-real-life-examples-of-machine-learning/#:~:text=Currently%2C%20machine%20learning%20has%20been,%2C%20learning%20association%2C%20regression%20etc.>> Citado na página 21.
- KISS, A. A. Heat-integrated reactive distillation process for synthesis of fatty esters. *Fuel Processing Technology*, Elsevier, v. 92, n. 7, p. 1288–1296, 2011. Citado 2 vezes nas páginas 13 e 24.
- KISS, A. A.; IGNAT, R. M. Enhanced methanol recovery and glycerol separation in biodiesel production—dwc makes it happen. *Applied Energy*, Elsevier, v. 99, p. 146–153, 2012. Citado 3 vezes nas páginas 14, 18 e 27.
- KRUKOWSKA, E.; AINGER, J. *EU Unveils Controversial Green Label for Gas and Nuclear*. 2022. Disponível em: <<https://time.com/6144289/eu-controversial-green-label-gas-nuclear/>>. Citado na página 18.
- LE, J. *The 10 Algorithms Machine Learning Engineers Need to Know*. 2016. Disponível em: <<https://www.kdnuggets.com/2016/08/10-algorithms-machine-learning-engineers.html>>. Citado 3 vezes nas páginas 22, 23 e 26.
- LENARDÃO, E. J. et al. "green chemistry": os 12 princípios da química verde e sua inserção nas atividades de ensino e pesquisa. *Química Nova*, SciELO Brasil, v. 26, p. 123–129, 2003. Citado na página 18.
- LEÃO, R. *Com aumento da disponibilidade, uso do gás natural é desafio estratégico*. 2020. Disponível em: <<https://editorabrasilenergia.com.br/com-aumento-da-disponibilidade-uso-do-gas-natural-e-desafio-estrategico/#:~:text=Segundo%20estimativas%20da%20Empresa%20Brasileira,o%20escoamento%20do%20g%C3%A1s%20natural.>> Citado 2 vezes nas páginas 14 e 18.
- LITTLE, A. D.; WALKER, W. H.; LEWIS, W. K. *The notion of a new kind of engineer—a chemical engineer—who understood both chemical processes and mechanical equipment was broached in England around 1880*. 1880. Disponível em: <<https://www.sciencehistory.org/historical-profile/arthur-d-little-william-h-walker-and-warren-k-lewis>>. Citado na página 16.
- LOURENÇO, L. A. Ciência de dados na engenharia química: Um estudo bibliométrico. *Congresso Brasileiro Interdisciplinar em Ciência e Tecnologia*, CoBICET, v. 1, n. 1, 2021. Citado na página 24.
- MAAOUANE, M. et al. Modelling industry energy demand using multiple linear regression analysis based on consumed quantity of goods. *Energy*, Elsevier, v. 225, p. 120270, 2021. Citado 2 vezes nas páginas 24 e 26.
- MACMURRAY, J. C.; HIMMELBLAU, D. Modeling and control of a packed distillation column using artificial neural networks. *Computers & chemical engineering*, Elsevier, v. 19, n. 10, p. 1077–1088, 1995. Citado na página 24.
- MATZOPOULOS, M. Dynamic process modeling: Combining models and experimental data to solve industrial problems. *Process systems engineering*, v. 7, p. 3–33, 2011. Citado na página 20.

- MCCABE, W. L.; SMITH, J. C.; HARRIOTT, P. *Unit operations of chemical engineering*. [S.l.]: McGraw-hill New York, 1993. v. 5. Citado na página 20.
- MOTA, C. J.; SILVA, C. X. da; GONÇALVES, V. L. Gliceroquímica: novos produtos e processos a partir da glicerina de produção de biodiesel. *Química Nova*, SciELO Brasil, v. 32, n. 3, p. 639–648, 2009. Citado 2 vezes nas páginas 17 e 18.
- MURTI, P.; WINKLE, M. V. Vapor-liquid equilibria for binary systems of methanol, ethyl alcohol, 1-propanol, and 2-propanol with ethyl acetate and 1-propanol-water. *Industrial & Engineering Chemistry Chemical and Engineering Data Series*, ACS Publications, v. 3, n. 1, p. 72–81, 1958. Citado na página 19.
- NARCISO, D. A.; MARTINS, F. Application of machine learning tools for energy efficiency in industry: A review. *Energy Reports*, Elsevier, v. 6, p. 1181–1199, 2020. Citado na página 21.
- PEDAMKAR, P. *Uses of Machine Learning*. 2022. Disponível em: <<https://www.educba.com/uses-of-machine-learning/>>. Citado na página 21.
- PEREIRA, F. S. G. et al. Produção de biodiesel metílico com óleo purificado de moringa oleifera lamarck. *Revista Virtual de Química*, v. 8, n. 3, p. 873–888, 2016. Citado na página 18.
- PERLINGEIRO, C. A. G. *Engenharia de processos: análise, simulação, otimização e síntese de processos químicos*. [S.l.]: Editora Blucher, 2005. Citado na página 19.
- PHELPS, F. A practical guide to chemical process optimization: Analysis of a styrene plant. 2016. Citado na página 20.
- QUISPE, C. A.; CORONADO, C. J.; JR, J. A. C. Glycerol: Production, consumption, prices, characterization and new trends in combustion. *Renewable and sustainable energy reviews*, Elsevier, v. 27, p. 475–493, 2013. Citado 2 vezes nas páginas 14 e 17.
- ROSCHER, R. et al. Explainable machine learning for scientific insights and discoveries. *Ieee Access*, IEEE, v. 8, p. 42200–42216, 2020. Citado na página 22.
- SANCHIS-SEBASTIÁ, M. et al. Techno-economic evaluation of biorefineries based on low-value feedstocks using the biosteam software: A case study for animal bedding. *Processes*, Multidisciplinary Digital Publishing Institute, v. 8, n. 8, p. 904, 2020. Citado na página 28.
- SCHNEIDER, L. et al. Otimização do processo de destilação em planta piloto para produção de bioetanol hidratado. In: *CONGRESSO BRASILEIRO DE ENGENHARIA QUÍMICA*. [S.l.: s.n.], 2014. v. 10. Citado na página 20.
- SCHWEIDTMANN, A. M. et al. Machine learning in chemical engineering: A perspective. *Chemie Ingenieur Technik*, Wiley Online Library, v. 93, n. 12, p. 2029–2039, 2021. Citado na página 21.
- SHI, R.; GUEST, J. S. Biosteam-lca: an integrated modeling framework for agile life cycle assessment of biorefineries under uncertainty. *ACS Sustainable Chemistry & Engineering*, ACS Publications, v. 8, n. 51, p. 18903–18914, 2020. Citado na página 28.
- SINGH, V.; GUPTA, I.; GUPTA, H. Ann based estimator for distillation—inferential control. *Chemical Engineering and Processing: Process Intensification*, Elsevier, v. 44, n. 7, p. 785–795, 2005. Citado na página 24.

SINGH, V.; GUPTA, I.; GUPTA, H. Ann-based estimator for distillation using levenberg–marquardt approach. *Engineering Applications of Artificial Intelligence*, Elsevier, v. 20, n. 2, p. 249–259, 2007. Citado na página 24.

SONDH, K. *In the 5th Industrial Revolution, creativity must meet technology*. 2021. Disponível em: <<http://blog.oxfordeconomics.com/world-post-covid/in-the-5th-industrial-revolution-creativity-must-meet-technology>>. Citado na página 21.

STANKIEWICZ, A.; MOULIJN, J. A. Process intensification. *Industrial & Engineering Chemistry Research*, ACS Publications, v. 41, n. 8, p. 1920–1924, 2002. Citado na página 17.

TEIXEIRA, A. C. Inferências em coluna de destilação multicomponente. *Monografia. Curso de Graduação em Engenharia Química*, 2003. Citado na página 20.

THANGARAJ, J.; NARAYANAN, R. L. Industry 1.0 to 4.0: The evolution of smart factories. 10 2018. Citado na página 22.

TURTON, R. et al. *Analysis, synthesis and design of chemical processes*. [S.l.]: Pearson Education, 2008. Citado na página 16.

UMPIERRE, A.; MACHADO, F. Gliceroquímica e valorização do glicerol. *Revista Virtual de Química*, v. 5, n. 1, p. 106–116, 2013. Citado na página 17.

VIANA, M. B. *Produção de biogás a partir de glicerol oriundo de biodiesel*. Tese (Doutorado) — Universidade de São Paulo, 2011. Citado na página 17.

WEICHERT, D. et al. A review of machine learning for the optimization of production processes. *The International Journal of Advanced Manufacturing Technology*, Springer, v. 104, n. 5, p. 1889–1902, 2019. Citado na página 13.

YUFENG, G. *The 7 Steps of Machine Learning*. 2017. Disponível em: <<https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e>>. Citado 2 vezes nas páginas 21 e 59.

ZHANG, J.; LIANG, S.; FENG, X. A novel multi-effect methanol distillation process. *Chemical Engineering and Processing: Process Intensification*, Elsevier, v. 49, n. 10, p. 1031–1037, 2010. Citado na página 28.

Apêndices

APÊNDICE A – CÓDIGO DE SIMULAÇÃO DA COLUNA DE DESTILAÇÃO COM MÓDULO BIOSTEAM

```
MethanolDistillation.py x
MethanolDistillation.py > generate_results
1 # Importação da classe Destilação Binária e variáveis de configuração dela:
2 from biosteam.units import BinaryDistillation
3 from biosteam import Stream, settings
4
5 # Função que gera os resultados da destilação:
6 def generate_results(water, methanol, glycerol, temp, k_value, pressure):
7     # Define o pacote de propriedades:
8     settings.set_thermo(["Water", "Methanol", "Glycerol"], cache=True)
9
10    # Define as propriedades da alimentação:
11    feed = Stream("feed", flow=(water, methanol, glycerol))
12    feed.T = temp
13
14    # Cria a destilação, simula e retorna seus custos de utilidade e projeto:
15    D1 = BinaryDistillation(
16        "D1",
17        ins=feed,
18        outs=("distillate", "bottoms product"),
19        LHK=("Methanol", "Water"), # Componentes leve e pesado
20        y_top=0.99, # Fração do componente leve no destilado
21        x_bot=0.01, # Fração do componente pesado no produto de fundo
22        k=k_value, # Razão de refluxo sobre o refluxo mínimo
23        is_divided=True, # Se as sessões de retificação e stripping são separadas
24        P=pressure, # Pressão de operação
25    )
26    D1.simulate()
27    # Etapa que realiza a soma dos custos de projeto para retornar o total:
28    total_purchase_cost = 0
29    for cost in D1.purchase_costs:
30        total_purchase_cost += D1.purchase_costs[cost]
31    return (D1.utility_cost, total_purchase_cost)
32
```

APÊNDICE B – CÓDIGO COM FUNÇÕES DE INSERÇÃO DE DADOS E REGISTRO EM ARQUIVO CSV

```
values.py x
values.py > ...
1 # Função que escolhe aleatoriamente um valor em uma lista:
2 from secrets import choice
3
4 # Função que cria uma lista a partir de uma faixa de valores:
5 from numpy import arange
6
7 # Pacote para manipular arquivos CSV:
8 import csv
9
10 from MethanolDistillation import generate_results
11
12 # Gerador de valores de entrada para a simulação do destilador:
13 def generate_entry(start, stop, step):
14     __array = arange(start, stop, step)
15     __list = list(__array)
16     __one_value = choice(__list)
17     return round(__one_value, 1)
18
19
20 # Gerador de linhas completas com entrada e saída de dados da simulação do destilador:
21 def generate_row():
22     methanol_feed = generate_entry(30, 150, 5)
23     water_feed = round(methanol_feed * 0.114)
24     glycerol_feed = methanol_feed
25     k_value = generate_entry(1.1, 7, 0.1)
26     temp_value = generate_entry(40, 88, 2)
27     P_value = generate_entry(101325, 200000, 1000)
28     results = generate_results(
29         water_feed,
30         methanol_feed,
31         glycerol_feed,
32         temp_value
```

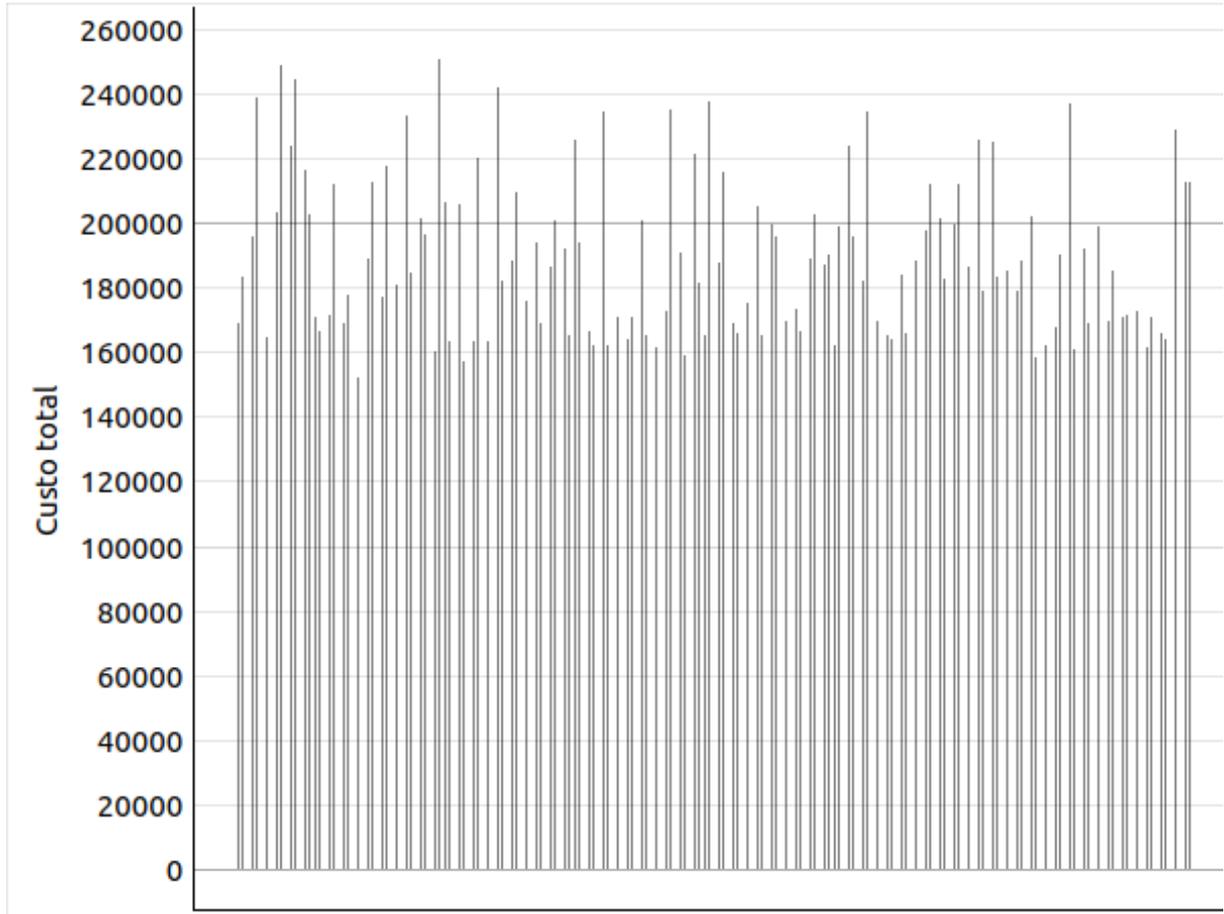
```
values.py x
values.py > ...
32     temp_value,
33     k_value,
34     P_value,
35 )
36     _row = (
37         k_value,
38         water_feed,
39         methanol_feed,
40         glycerol_feed,
41         temp_value,
42         P_value,
43         results[0],
44         results[1],
45     )
46     return __row
47
48
49 # Função que preenche o arquivo com os dados gerados na simulação:
50 def populate_file(path):
51     file = open(path, "w")
52     writer = csv.writer(file)
53     header = (
54         "RR",
55         "water_feed(kmol/hr)",
56         "methanol_feed(kmol/hr)",
57         "glycerol_feed(kmol/hr)",
58         "temperature(degC)",
59         "pressure(Pa)",
60         "utility_cost(USD/hr)",
61         "total_purchase_cost(USD)",
62     )
63     writer.writerow(header)
```

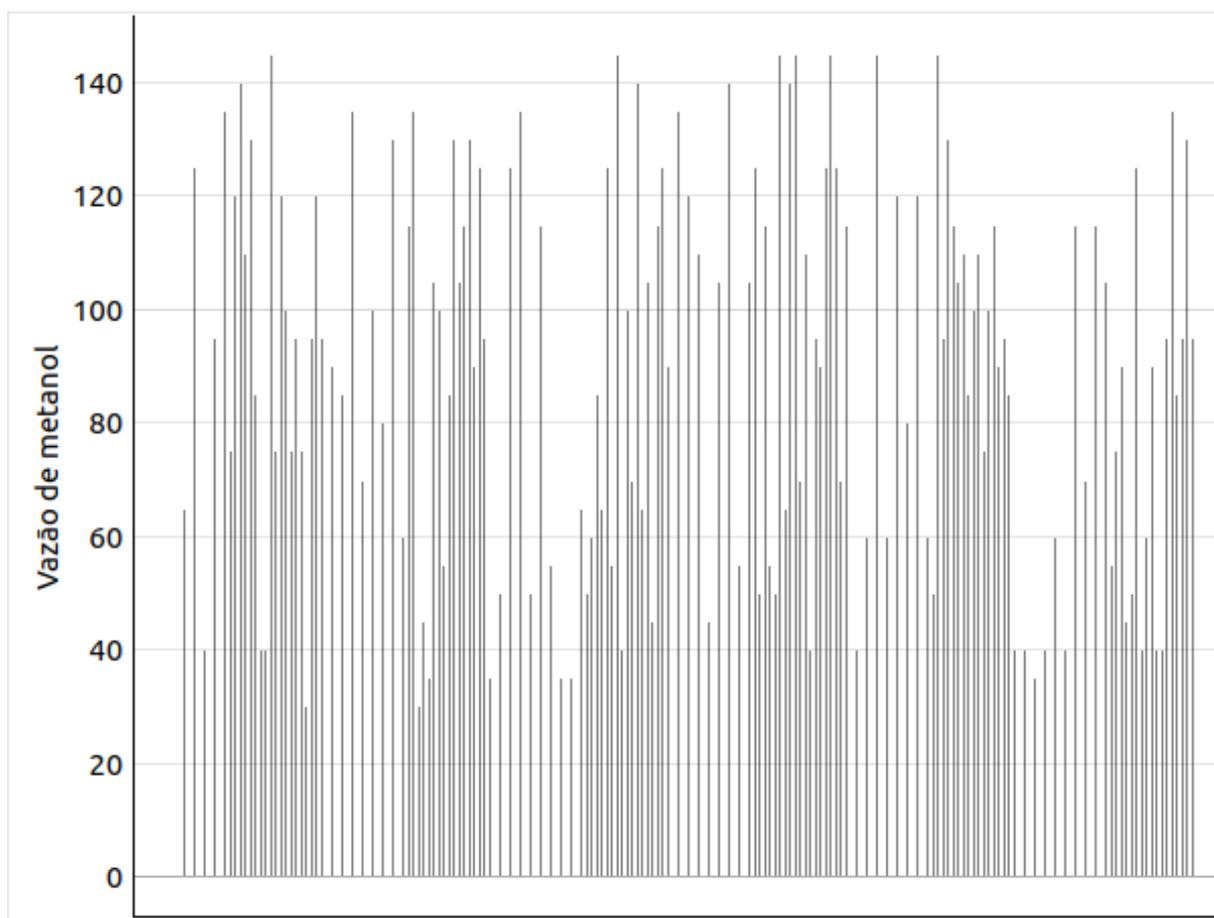
```
values.py x
values.py > ...
61         "total_purchase_cost(USD)",
62     )
63     writer.writerow(header)
64     value = 0
65     # Laço que adiciona linhas de dados ao arquivo na quantidade especificada:
66     while value in range(0, 1000):
67         writer.writerow(generate_row())
68         value += 1
69
70     file.close()
71
72
73 # Chamada da função com o caminho para o arquivo que deve se preenchido:
74 populate_file("./ins_data.csv")
75
```

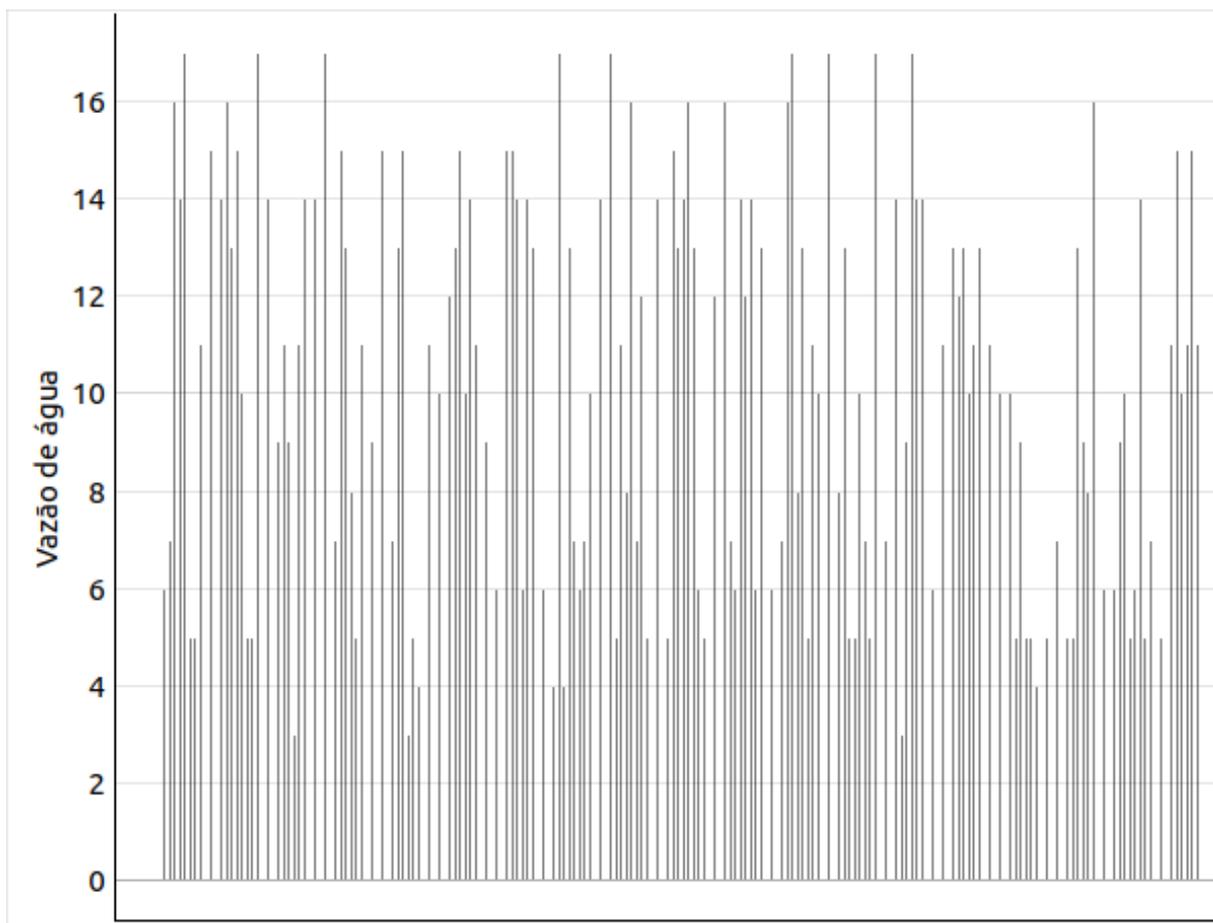
APÊNDICE C – CÓDIGO DO ARQUIVO DE DADOS CSV

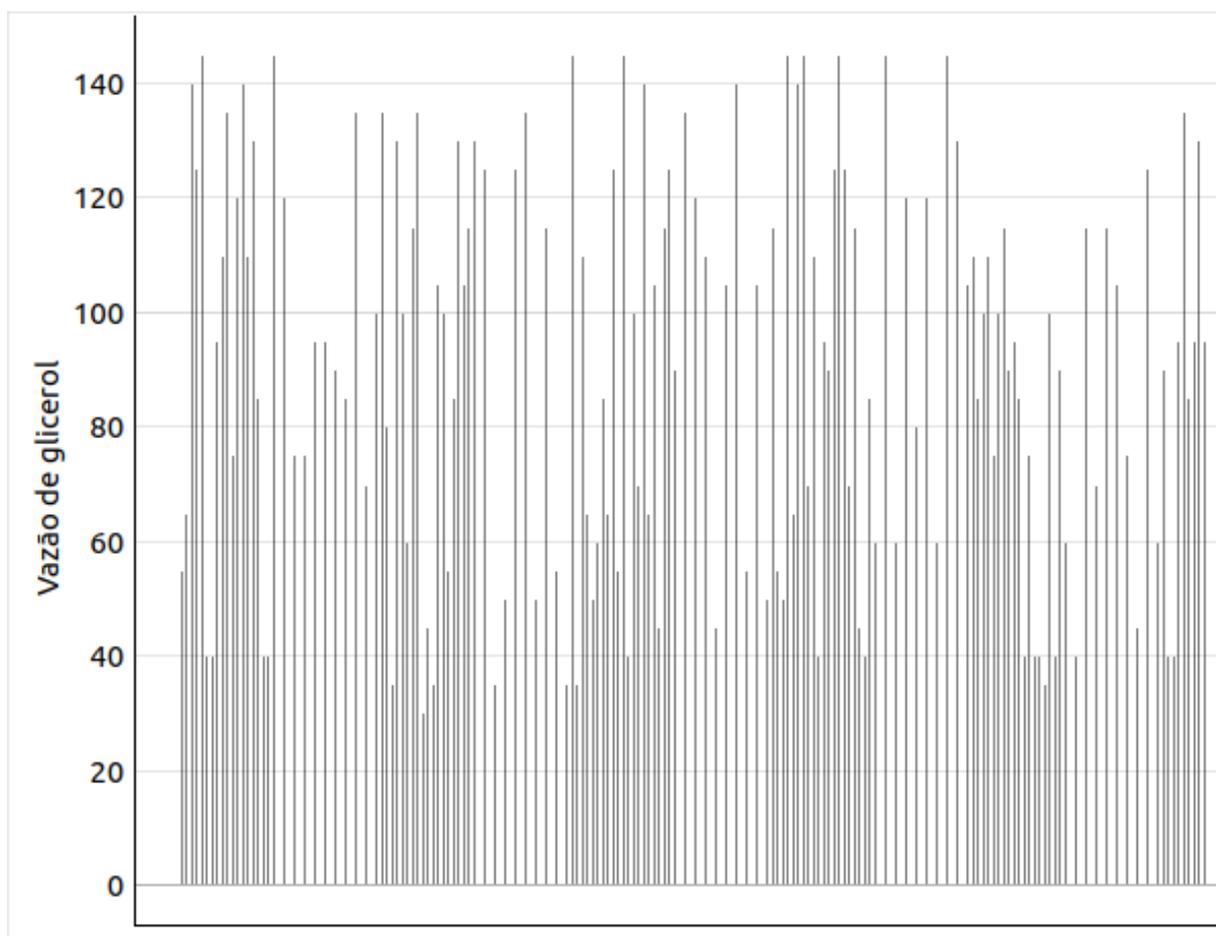
```
ins_data.csv
1 RR,water_feed(kmol/hr),methanol_feed(kmol/hr),glycerol_feed(kmol/hr),temperature(degC),pressure(Pa),utility_cost(USD/hr),total_purc
2 2.2,5,40,40,80,178325,60.72219409444035,164293.14422315653
3 2.1,8,70,70,44,152325,105.86567479346064,170101.7164604328
4 1.9,15,130,130,64,165325,190.8377664868415,188880.05040525066
5 2.8,14,125,125,62,194325,286.81804452676727,191132.43737431295
6 6.7,7,65,65,42,184325,210.9827346831329,183632.45435314445
7 1.5,11,95,95,80,158325,128.49419838154085,176293.63886784547
8 1.6,9,80,80,56,178325,115.53439593781738,173455.82346666558
9 1.9,7,60,60,80,159325,86.08384042547863,166466.52260915548
10 1.4,17,145,145,48,195325,208.8091590456094,199758.49578123403
11 2.9,13,110,110,48,149325,180.43274433663314,191520.52900091096
12 1.9,15,130,130,82,163325,186.81965767935498,189070.70611719182
13 1.4,5,40,40,84,168325,54.07108135645644,168981.85576889385
14 4.2,16,140,140,76,141325,251.710667887972,220081.42577476558
15 3.0,5,45,45,44,124325,72.25428608620017,159121.94797248536
16 1.6,10,90,90,80,125325,119.82638896681331,171670.1602621873
17 2.6,7,60,60,54,160325,95.89396548633009,164434.42391856402
18 1.9,16,140,140,44,166325,209.8398003390944,193179.07540003146
19 1.4,11,95,95,76,166325,129.44817468241538,176000.96103994298
20 4.5,15,135,135,78,164325,255.86734302125848,221765.6250762072
21 5.6,6,55,55,68,110325,106.20404063776014,170638.8012580719
22 6.8,6,55,55,72,101325,113.80698002101200,180542.08155322552
23 2.7,7,65,65,64,166325,104.33329810839224,166398.47884400855
24 6.8,16,140,140,56,189325,336.7006288593072,262715.28036688385
25 6.2,15,130,130,60,170325,291.5810969627031,242565.86600103584
26 2.4,4,35,35,58,110325,51.247445954213156,162487.85715255828
27 4.6,9,75,75,78,139325,139.93785398224648,184207.08762315375
28 2.6,9,75,75,54,174325,121.71708660740775,169271.81849382745
29 6.5,14,125,125,42,179325,292.0139690269742,243774.45468855932
30 2.8,11,95,95,44,195325,221.89981632794021,175991.5808733529
31 6.1,7,65,65,44,134325,137.77291468752273,182068.8050063193
32 1.5,12,105,105,78,176325,145.1363760201628,182303.06882525087
```

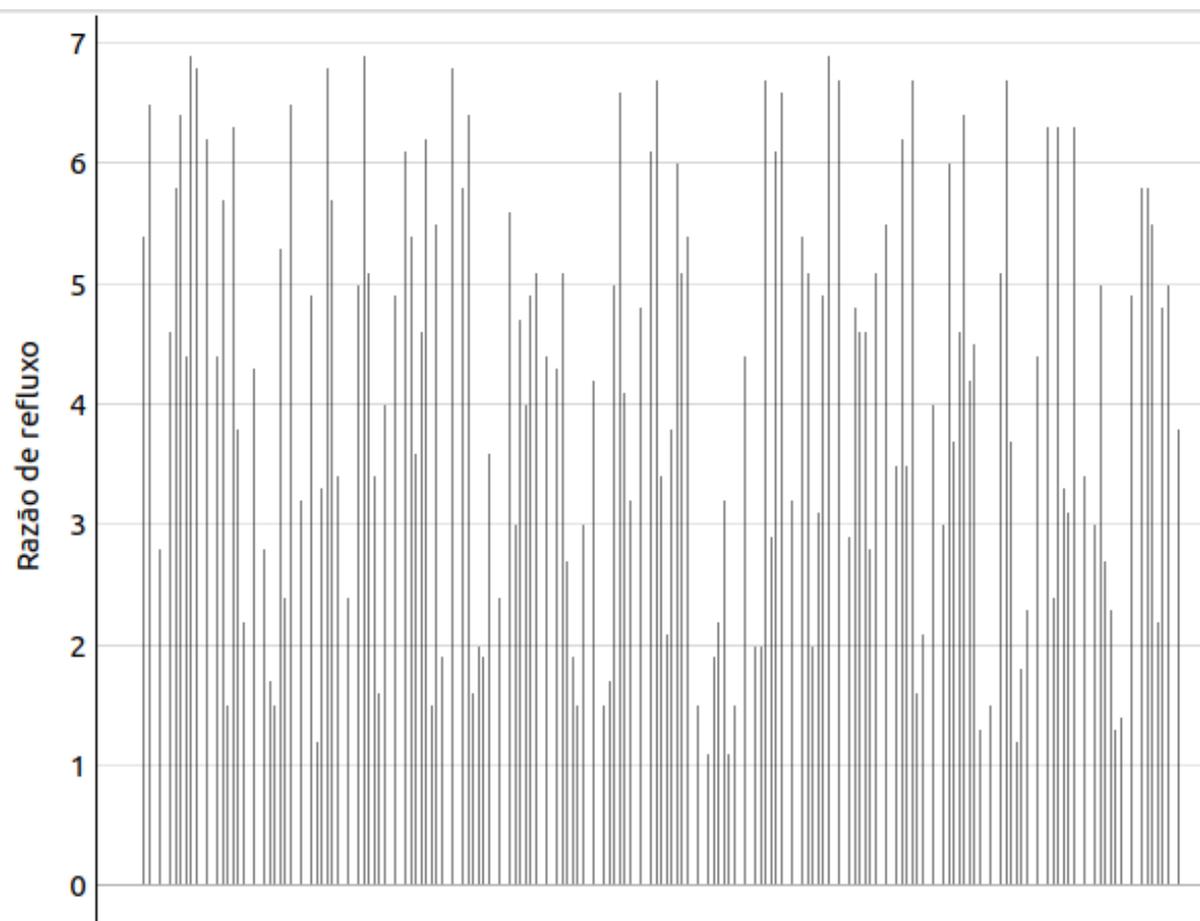
APÊNDICE D – HISTOGRAMAS DE DISTRIBUIÇÃO DOS DADOS SIMULADOS

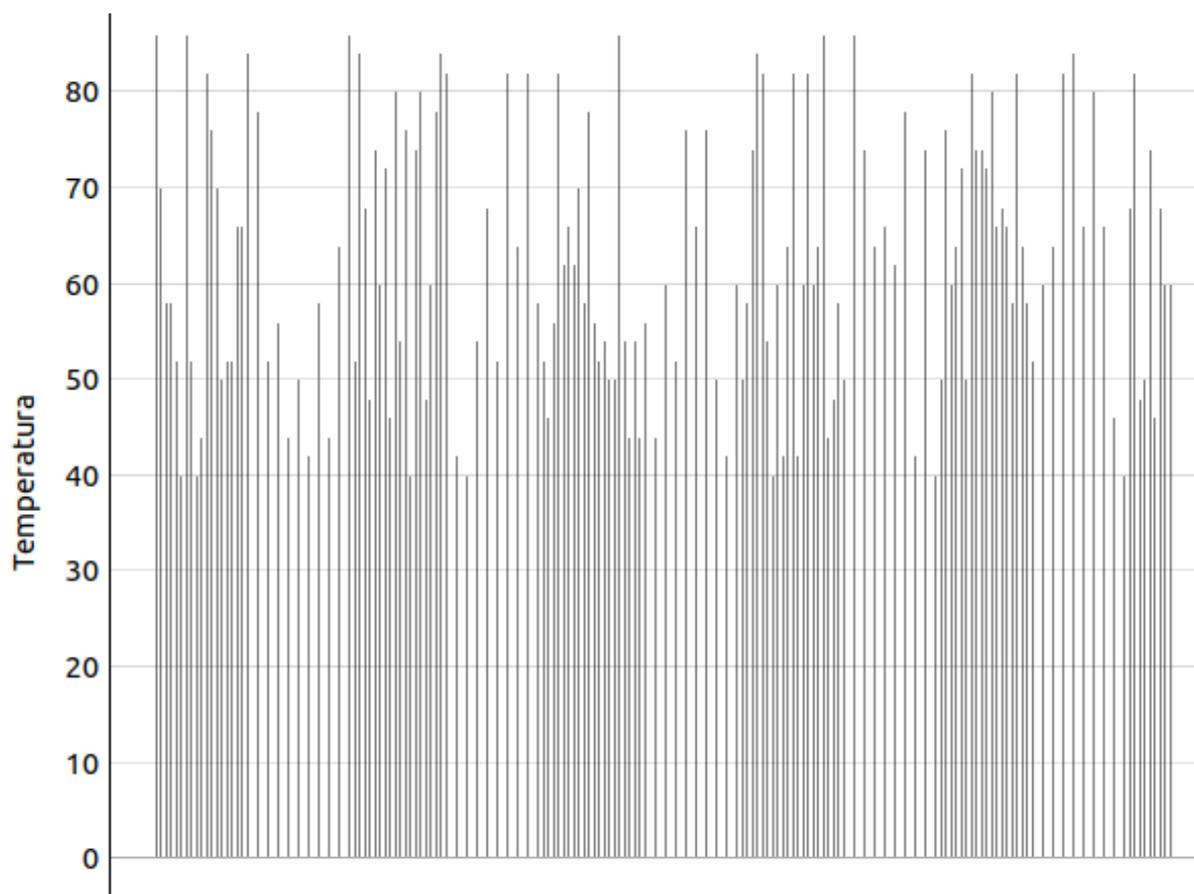


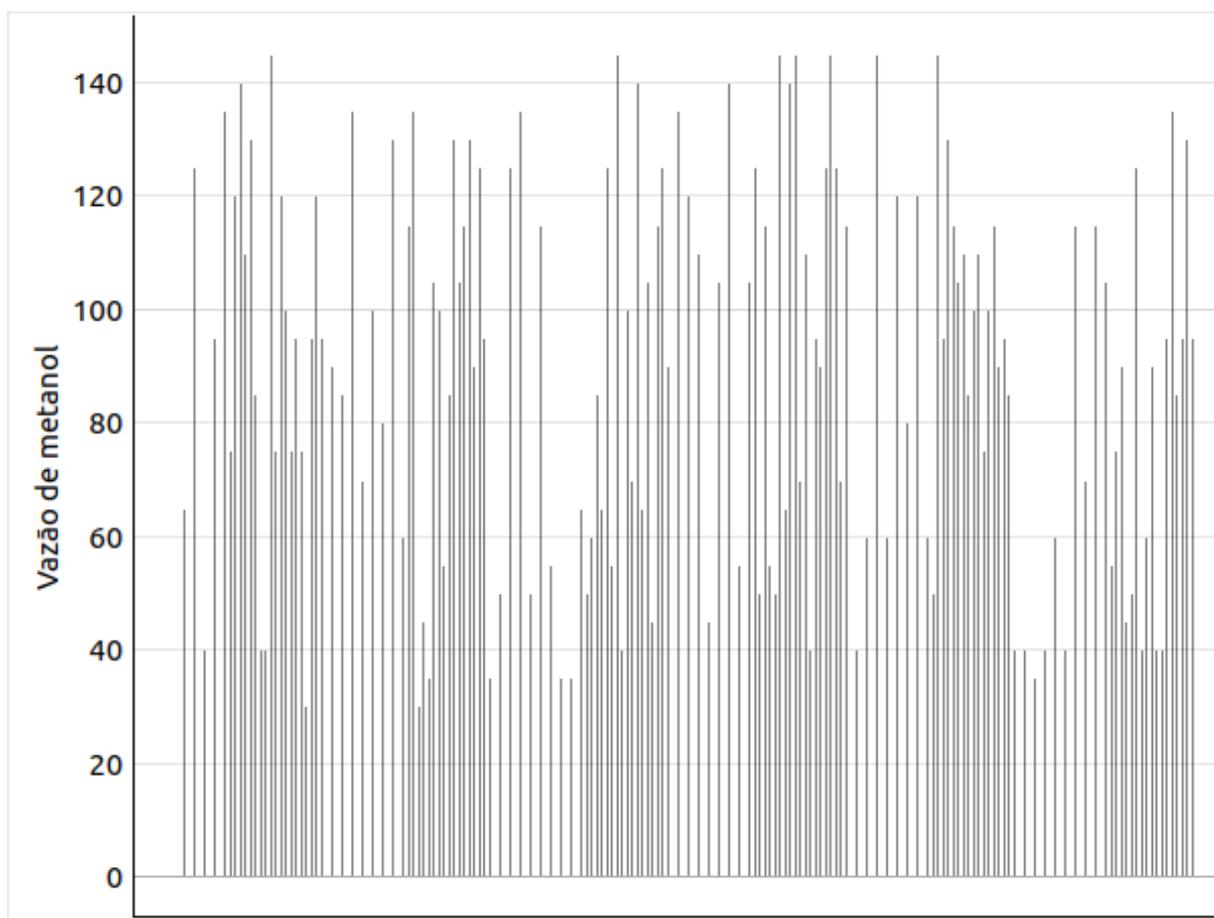








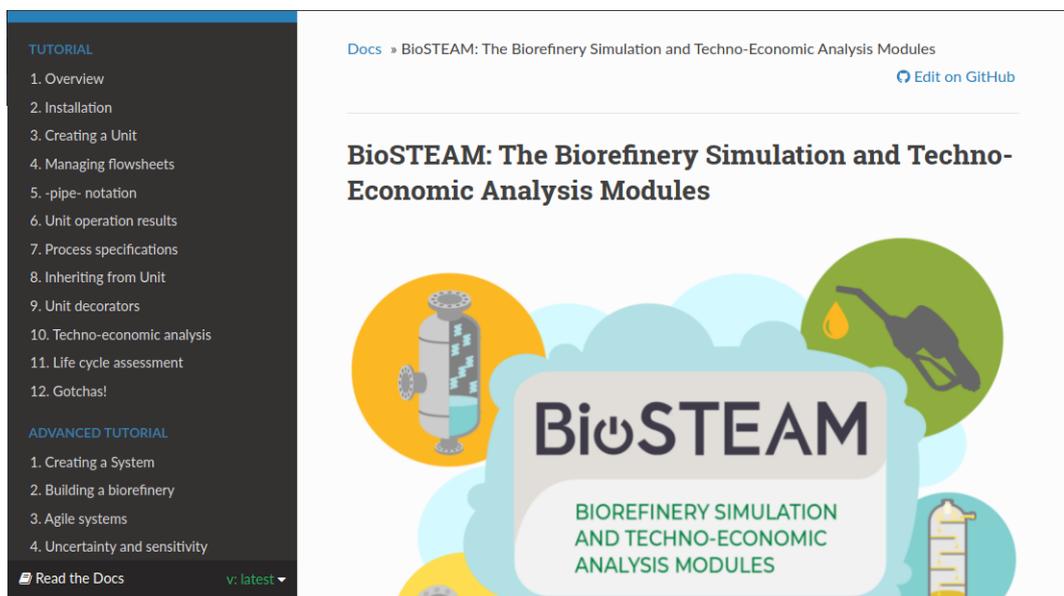




Anexos

ANEXO A – SIMULADOR BIOSTEAM: MÓDULOS DE SIMULAÇÃO DE BIORREFINARIAS E ANÁLISES TECNO-ECONÔMICAS

Este simulador pode ser acessado em <<https://biosteam.readthedocs.io/en/latest/>>, ele oferece um tutorial didático de utilização dos módulos, além de uma grande biblioteca de exemplos de unidades industriais e sistemas auxiliares.



Possui um repositório público no endereço:
<<https://github.com/BioSTEAMDevelopmentGroup/biosteam>>

ANEXO B – FERRAMENTAS PARA INSERÇÃO DE DADOS NO SIMULADOR, REGISTRO DE RESULTADOS OBTIDOS E GERAÇÃO DE ARQUIVO DE TRABALHO

A linguagem Python de programação é a mais indicada para algoritmos de aprendizado de máquina (YUFENG, 2017) e sua documentação pode ser acessada em:

<<https://docs.python.org/3/>>

As bibliotecas Python "secrets", "numpy" e "csv" tem suas documentações disponíveis respectivamente em:

secrets

<<https://docs.python.org/3/library/secrets.html>>

numpy

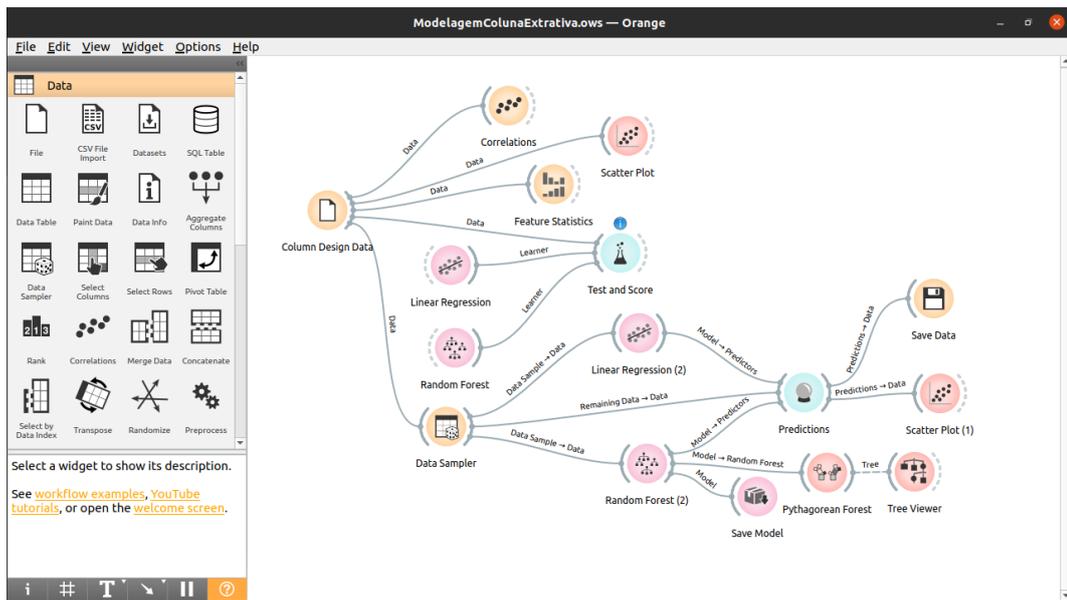
<<https://numpy.org/doc/>>

csv

<<https://docs.python.org/3/library/csv.html>>

ANEXO C – SOFTWARE ORANGE DATA MINING PARA AVALIAÇÃO, TREINAMENTO E TESTES DE MODELOS DE APRENDIZADO DE MÁQUINA

O software Orange Data Mining possui uma interface gráfica extremamente intuitiva, baseada na conexão de widgets e ferramentas que vão muito além do escopo deste trabalho:



Sua documentação pode ser acessada no endereço:
<<https://orangedatamining.com/docs/>>