



Desenvolvimento de um banco de dados com informações de processos de mistura e reação química em microdispositivo

Aluno: Wilson Legnaro Volpi Nakano
Orientador: Prof. Dr. Osvaldir Pereira Taranto
Co-orientador: Dr. Harrson Silva Santana

Resumo

Comumente utilizado em distintas áreas, os bancos de dados são ferramentas ordenam e coordenam dados dinamizando as pesquisas científicas. Tendo em vista que um grande número de fatores influencia processos químicos em dispositivos microfluídicos e nem sempre todas as informações do processo estão disponíveis aos pesquisadores, o projeto teve como proposta o desenvolvimento de um banco de dados em que se armazene dados referentes à processos químicos em microrreatores | previamente coletados. O modelo optado baseou-se na utilização de banco de dados relacional PostgreSQL e a criação de uma interface de usuário utilizando a linguagem programacional Python. A aplicabilidade do programa elaborado foi testada através de inserção e coleta de dados, no qual, posteriormente foram manipulados em métodos de regressão linear utilizando estruturas de redes neurais, como é o caso das estruturas de machine learning.

Palavras-chave:

Microfluídica, Banco de dados, Inteligência Artificial.

1. Introdução

Dentro da ótica da expansão da indústria 4.0 na sua concretude, a geração de um enorme montante de dados é recorrente à sociedade, o termo *Big Data* passou a referenciar esse volume de forma genérica e toda vez que seu volume chega a valores críticos, necessita-se desenvolver novas tecnologias capazes de armazenar, processar e alterando mediante a evolução da capacidade de hardwares.

Dependendo da quantidade de dados que estão sendo gerados é necessário um banco de dados que é basicamente um conjunto de dados que armazena informações em formato sequencial ou não-sequencial.

Os dados são uma coleção de *bits* e *bytes* de informação e um banco de dados é uma coleção organizada de dados. Do lado mais amplo, um sistema de gerenciamento de banco de dados (DBMS) é um conjunto de *softwares* responsáveis pelo gerenciamento de um banco de dados, ou seja, é um *software* computacional que interage com o usuário, outros aplicativos e o próprio banco de dados para capturar e analisar dados (Choudhary et al., 2016). Assim, percebe-se que para analisar dados, ou extrair algum conhecimento de um conjunto de informações são necessários os dados propriamente, o gerenciamento dos dados, comunicação da aplicação ou programa com o DBMS e posteriormente a comunicação entre o sistema operacional e do hardware onde os dados estão armazenados (Figura 1) (Foster et al., 2016).

Tecnicamente, um sistema de banco de dados (*database system*, “DBS”) é um sistema informatizado de manutenção de

registros com a proposta de manter informações e de torná-las disponíveis sempre que requeridas (Foster et al., 2016). Para a formulação de sistemas de banco de dados, é preciso adequar a construção e a arquitetura a 3 pontos intrínsecos: confiabilidade, escalabilidade e manutenibilidade (Kleppmann et al., 2016).

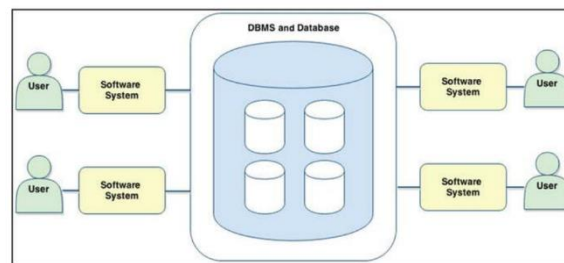


Figura 1: Banco de dados abordado como gerenciamento de dados (Foster et al., 2016)

Para a utilização do banco de dados, requer-se a existência de um sistema de gerenciamento do banco de dados (*database management system*), DBMS. Como um dos modelos mais utilizados, o usuário comunica-se através de uma aplicação ou programa (interface) com o DBMS. Posteriormente há o diálogo entre sistema operacional e *hardware* onde os dados estão armazenados (Foster et al., 2016).

Ao passo que se tem consolidação do banco de dados visa-se a utilização dele ao propósito que motivou sua concepção. Considerando-se que o projeto em questão tem em vista a feitura de um *software* suporte à uma área da engenharia química, tal qual o é o estudo da microfluídica, é necessário tomar como referência ao

projeto atender as demandas da área ou principais utilizações. Elucida-se quais são as áreas que os dados e o banco de dados são utilizados como ferramenta: simulações de reações químicas e processos de mistura, desenvolvimento de estruturas de microrreatores, e utilização de modelos de regressão com inteligência artificial (redes neurais artificiais e variações).

No que tange o assunto visado no terceiro tópico acima, “Utilização de modelos de regressão com inteligência artificial (redes neurais artificiais e variações)”, verifica-se uma ampla e crescente utilização de tais métodos dentro da comunidade científica.

Em Orimoto et al. (2012), utilizou-se um método baseado em redes neurais artificiais para analisar um total de 3404 conjuntos de dados de sínteses combinadas de nanopartículas NPs de CdSe com seis parâmetros para adquirir as relações de condição-propriedade nos conjuntos de dados. Segundo os autores, a síntese de nanomateriais é extremamente sensível a vários fatores das condições experimentais. Portanto, para controlar a síntese, é importante averiguar de forma abrangente as relações entre as condições e as propriedades dos nanomateriais. Os dados utilizados pelos pesquisadores foram obtidos usando sínteses combinadas de nanopartículas-CdSe em microrreatores.

Em contra partida aos dados utilizados no caso anterior, Moayedi et al. (2020) através de 30 conjuntos de dados de tipo de óleo, catalisador, concentração de catalisador, temperatura de reação, razão de alimentação entre metanol e óleo, tempo de reação, frequência e amplitude de ondas supersônicas, aplicou-se quatro distintos modelos de aprendizado de máquina (*Machine Learning*) – Alternating Model Tree (AMT), Random Tree (RT), Least Median Square Regression (LMSR) e Multiple Linear Regression (MLR) – para a predição de rendimento (pureza) no processo de produção de biodiesel; 80 por cento dos dados foram utilizados para treinamento dos modelos e os demais para a valoração. A partir dos resultados obtidos, expõem-se a aplicabilidade dos métodos AMT, RT e MLPR, dada que os valores de erros absolutos calculados mostram respectiva veracidade; ao método LMSR, por não apresentar essa mesma correlação com dados experimentais, não se pode confirmar sua aplicabilidade. Segundo R. Lo (2013) a “Microfluídica” é definida como a ciência e tecnologia que envolve o estudo do comportamento dos fluídos, manipulação controlada de fluídos e o *design* de dispositivos ou sistemas que possa executar de forma confiável tarefas em microcanais com dimensões típicas de dezenas a centenas de micrômetros. As ferramentas desenvolvidas nessa área possibilitaram obter grandes vantagens nos processos em microescala, comparados aos métodos macroscópicos. As principais características dos sistemas microfluídicos são: menor quantidade de reagentes e amostras; tempo de reações menores; baixo custo de fabricação; área superficial elevada em relação ao volume (10000 m² m⁻³) e elevada transferência de massa e calor (Fogler, 2006; Whitesides 2006). Suas aplicações compreendem desde sistemas biológicos, microsensores, microrreatores a microtrocaadores de calor (Skardal et al., 2017; Cardoso et al. 2017; Santana et al. 2016; Santana et al. 2017). A obtenção de altas conversões e seletividade aliada a curtos tempos de residência nesses microsistemas estão possibilitando tornar os processos mais eficientes, o que resulta na criação de novos produtos e métodos de produção de forma mais rápida para a indústria.

Nos últimos anos foram realizados diversos estudos sobre síntese química em microdispositivos ou microrreatores (Suryawanshi et al. 2018). Os estudos abordam diversas reações química, influência do tipo de geometria, os efeitos da variação da temperatura, concentração de catalisador e razão molar entre os reagentes, a relação entre o grau de mistura dos reagentes e eficiência reacional e o escalonamento de microunidades. Assim, se observa que um grande número de fatores influencia o processo de

síntese de biodiesel em microestruturas. Como foi visto, é possível utilizar essas microestruturas para uma produção industrial, mas para isso se tornar comum nas indústrias é necessário o *design* e otimização de processos e microestruturas levando em consideração todos os aspectos mencionados acima.

Exposto isso, sabendo que a microfluídica é uma nova ciência e sabendo da sua carência de dados, buscou-se criar um banco de dados num conforme teórico dos conceitos de um “sistema de gerenciamento de banco de dados” (DBMS) e um “sistema de banco de dados” (DBS). Para tal, criou-se um banco de dados relacionável em “PostgreSQL” e uma interface de usuário em “Python” (versões utilizadas serão detalhadas posteriormente). As diversas e possíveis formas de se salvar os dados, buscou-se priorizar um formato que facilitaria a reprodução dos dados à futuras utilizações. Por fim, como uma primeira tentativa, utilizou-se de recursos ligados à inteligência artificial para comprovar a eficácia do programa elaborado, nesse sentido, o escopo da pesquisa realizada através do projeto “Desenvolvimento de um banco de dados com informações de processos de mistura e reação química em microdispositivos” teve como premissa a abordagem de todos os temas discutidos até dado momento.

2. Materiais e Métodos

2.1 Coleta de dados:

Primeiramente foi realizada uma busca entre os diretórios de acervos científicos a serem utilizados; a lista de bancos de dados a serem utilizados gerada foi: *Web of Science, Scopus, Science Direct, Scielo, arXiv, Wiley Online Library, Springer, ProQuest Central, Emerald Insight, Taylor and Francis, ACS Publications, Periódicos CAPES e Industrial and Engineering Chemistry Research*. Dado a a maior presença de resultados nos acervos *Science Direct, Web of Science e Scopus* limitei a pesquisa a utilização destes.

Como um modelo de pesquisa à artigos acadêmicos, tomando como referência o modelo executado em Albahri et al. (2020), exemplifica-se o diagrama exposto (Figura 2).

2.2 Banco de dados:

Anterior ao desenvolvimento do banco de dados, utilizou-se planilhas do *Microsoft Excel (Microsoft Office 365)* para armazenar previamente os dados e informações coletados a respeito de processos de mistura e reação química em microdispositivos. Ao passo que o banco de dados esteve estabelecido, os dados foram migrados para a estrutura correta DBS desenvolvido.

A partir do conceito apresentado do “DBMS” a construção e estruturação do banco de dados optada consiste na utilização da linguagem *SQL* e ao banco de dados em *PostgreSQL*. A criação do banco foi realizada utilizando os softwares *DataGrip 2019.3.4* e *DBeaver 7.0.0.202003021717*.

2.3 GUI (Graphical User Interface):

Para o desenvolvimento do *Graphical User Interface* (interface de usuário) utilizou-se a linguagem de programação *Python* na versão 3.7.6. Como IDE (*Integrated Development Environment*) utilizou-se o software *PyCharm Professional Edition with Anaconda plugin 2019.3.4* através da plataforma *Anaconda* na versão 4.8.3.

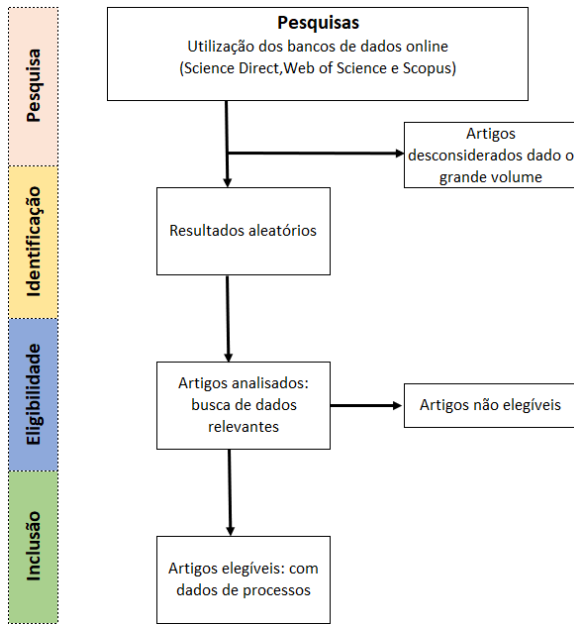


Figura 2: Esquema para coleta de dados em acervos científicos.

2.4 Aprendizado de máquina:

A formulação de métodos de inteligência artificial (*machine learning*) utilizou a ferramenta *Jupyter Notebook* na versão 6.0.3 através da plataforma *Anaconda* na versão 4.8.3. Utilizando a proposta em Moyaedi et al. (2020), os métodos de *machine learning* optados a serem utilizados foram *Least Median Squares Regression (LMSR)*, *Ridge Regression* e *K-Nearest Neighbors (KNN)*. Todos os conjuntos de dados utilizados foram segregados em duas parcelas: 80 por cento para treinamento das redes e 20 por cento para o teste das redes. Para a validação dos resultados, 4 índices estatísticos foram utilizados: R^2 , MAE (média absoluta do erro), MSE (erro quadrático médio) e RMSE (raiz quadrada do erro quadrático médio). As equações abaixo (1) - (4) descrevem as os índices R^2 , MAE, MSE e RMSE, respectivamente:

$$R^2 = 1 - \frac{\sum_{i=1}^s (Y_{ipred} - Y_{iobs})^2}{\sum_{i=1}^s (Y_{iobs} - \bar{Y}_{obs})^2} \quad (1)$$

$$MAE = \frac{1}{N} \sum_{i=1}^s |Y_{iobs} - Y_{ipred}| \quad (2)$$

$$MSE = \frac{1}{N} \sum_{i=1}^s (Y_{iobs} - Y_{ipred})^2 \quad (3)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^s [(Y_{iobs} - Y_{ipred})]^2} \quad (4)$$

Em todos os índices, a variável s denota o número de dados, as variáveis Y_{iobs} e Y_{ipred} representam os valores atuais (variável dependente) e preditos (predição da variável dependente), respectivamente; e a variável \bar{Y}_{obs} denota as médias dos valores atuais da variável dependente.

3. Resultados e Discussão

3.1 Coleta de dados:

Em tal ponto, representa-se novamente o diagrama que descreve o procedimento para coleta de dados em acervos científicos, adicionando-se as quantidades de artigos utilizadas para a consolidação do banco de dados, através da Figura 3.

A booliana utilizada na pesquisa, em conjunto com o retorno numérico de informações compatíveis, está descrito na Figura 4.

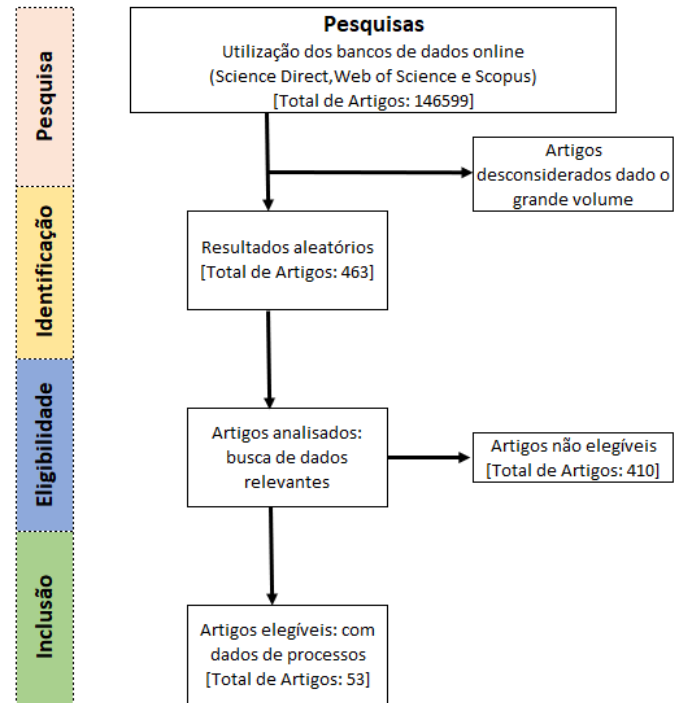


Figura 3: Esquema para coleta de dados em acervos científicos com quantidades de artigos encontradas.

Termos de detalhes de consulta	Resultados dos banco de dados
('microfluidic' OR 'micro fluidic' OR 'microreactor' OR 'micro reactor')&date=2015-2020	ScienceDirect = 72.538 artigos Web of Science = 31.813 artigos Scopus = 42.248 artigos
Total de artigos	146599

Figura 4: Booliana utilizada para pesquisa nos acervos científicos e resultados obtidos.

3.2 Banco de dados:

Através dos 53 artigos elegíveis com dados de processos, obteve-se 843 linhas de dados de processos que utilizam microrreatores. Após a consolidação do banco de dados no *PostgreSQL*, os dados foram migrados da planilha em *Excel* para o acervo, distribuindo-os de acordo com as 12 tabelas (Espécie Química, Processo, Microrreator, Processo de reação química, Processo de mistura, Misturador, Catalisador, Fluxo, Tempo de residência, Pressão, Temperatura e Constantes da cinética de reação).

3.3 GUI (Graphical User Interface):

A interface de usuário (GUI) criada apresenta 3 janelas de interação com o usuário: janela principal, janela para inserir dados ao banco de dados e janela para carregar dados do banco de dados.

Na janela principal encontra-se os logos da Faculdade de Engenharia Química e da UNICAMP, e informações estatísticas retiradas do banco de dados a respeito do número de processos armazenados separados por ano, os 5 processos com mais linhas de dados e por fim, os 5 microrreatores mais utilizados (Figura 5).

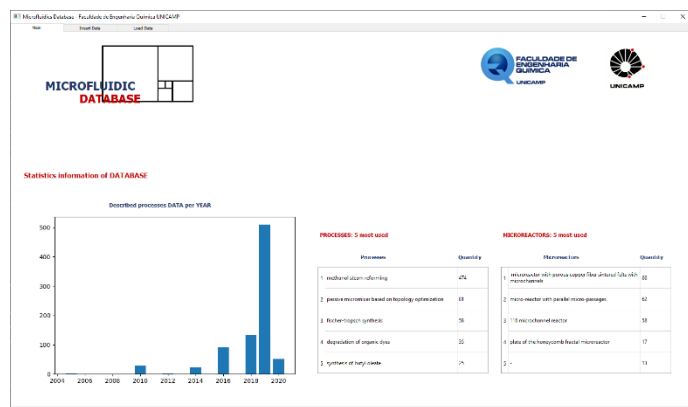


Figura 5: Janela principal da interface de usuário do banco de dados de microfluídica.

Na janela de inserção de dados, encontra-se meios para se importar arquivos nos formatos ".xlsx" e ".csv" e também a possibilidade de se escrever os dados manualmente. Posterior a descrição das informações, salva-se os dados por meio de um botão ao banco de dados em *PostgreSQL* (Figura 6).

Na janela de carregar dados, é possível optar por quais dados serão carregados por meio do processo descrito e do microrreator utilizado. Assim que determinado, o usuário pode exportá-los nos formatos ".xlsx" e ".csv" (Figura 7).

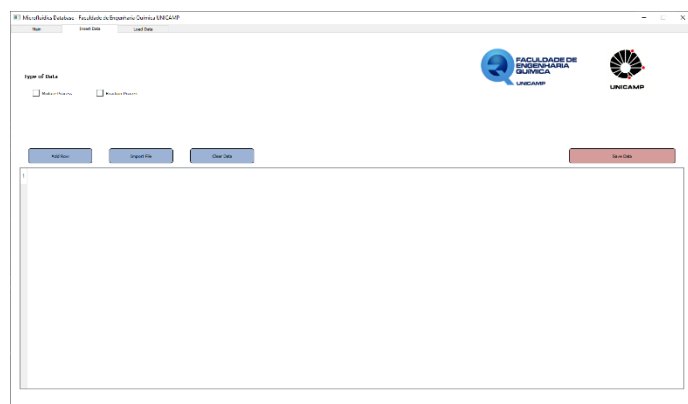


Figura 6: Janela de inserção de dados da interface de usuário do banco de dados de microfluídica.

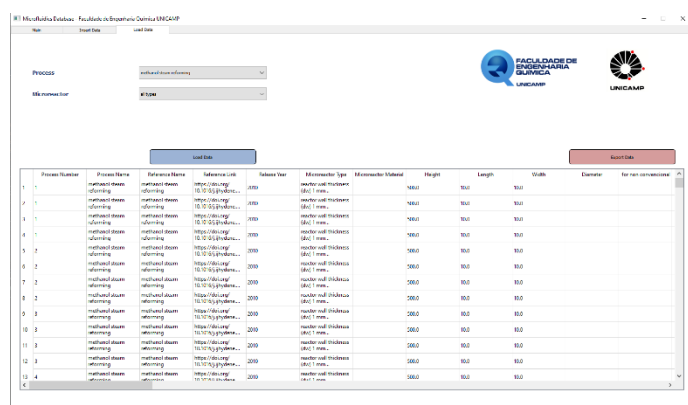


Figura 7: Janela de inserção de dados da interface de usuário do banco de dados de microfluídica.

3.4 Aprendizado de máquina:

Para a realização desta análise recorreu-se a dois conjuntos de dados coletados através do banco de dados desenvolvido. As informações são referentes aos processos "Reforma à vapor do

metanol" e "Otimização da topologia do micromisturador passivo", respectivamente condizentes à processo de reação química e processo de mistura. Para cada um dos métodos de machine learning realizou-se a validação através da comparação da variável dependente predita e observada (esperada). Para o caso descrito na Figura 8 a variável dependente refere-se ao rendimento na formação do produto (gás hidrogênio) devido a reação química e para o caso descrito na Figura 9 o índice de mistura.

Realiza-se a comparação dos índices estatísticos por meio da Figura 10 e Figura 11 (respectivamente às regressões: "Reforma à vapor de metanol" e "Otimização da topologia do micromisturador passivo". Infere-se que a partir dos valores obtidos para "R²" (menores que 0.95) nenhum dos modelos de rede apresentam uma acurácia elevada ao ponto de prever corretamente os rendimentos da reação química apresentada.

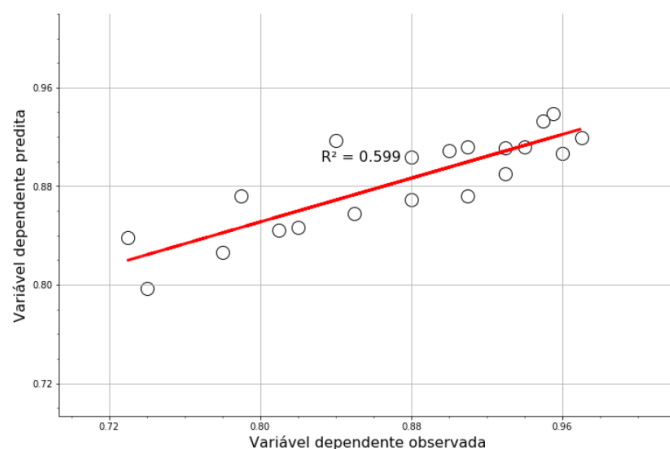


Figura 8: Output de rede para o conjunto de dados de teste (método Ridge).

Aponta-se como maior determinante para tal não racionalidade (entre os dados preditos e observados), a concentração de dados "rendimentos observado (esperada)" numa faixa pequena dentro da sua real variabilidade (entre 0 e 1); os dados utilizados situam-se na faixa de 0.78 a 1.0. Em segundo, o arranjo de dados compreendeu a análise de dados de reações químicas realizadas em distintos microrreatores; a continuidade do uso dos dados mesmo com tal afirmação foi motivada devido a uma baixa quantidade de dados em cada tipo específico de microrreator. Por último, utilizando-se um maior conjunto de dados seria possível verificar se alguma mudança ocorreu.

Infere-se, do mesmo princípio abordado ao caso da "Reforma à vapor de metanol", mas neste caso, "Otimização da topologia do micromisturador passivo", a não relacionabilidade da concentração de dados ocorre com o "índice de mistura observado (esperada)". A sua variabilidade também está compreendida na faixa entre 0 e 1. Também pela presença de diferentes estruturas de microrreator e, exclusiva a este conjunto de dados, a utilização de diferentes estruturas de micromisturadores. Para evitar tais resultados, é necessário o agregar o banco de dados com um maior número de informações deste processo de mistura.

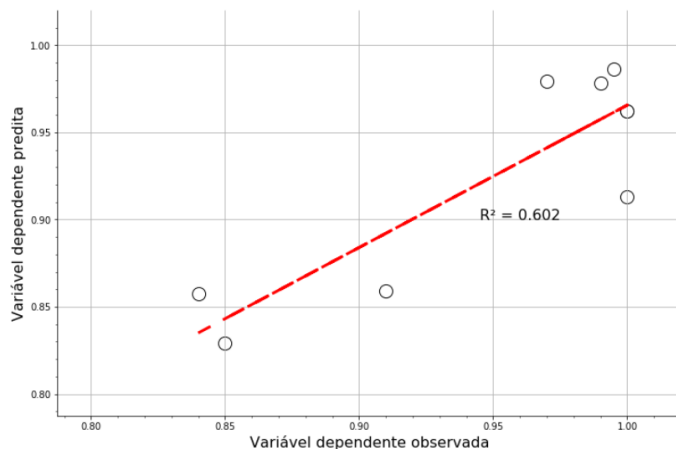


Figura 9: Output de rede para o conjunto de dados de teste (método KNN).

Modelos Proposto	Índices estatísticos			
	R ²	MAE	MSE	RMSE
Least Median Squares Regression (LMSR)	0.59450	0.02818	0.00135	0.03669
Ridge Regression	0.59948	0.03730	0.00213	0.04613
K-Nearest Neighbors (KNN)	0.44760	0.04210	0.00218	0.04666

Figura 10: Resultado das redes propostas ao conjunto de dados “Reforma à vapor do metanol”, considerando os índices estatísticos.

Modelos Proposto	Índices estatísticos			
	R ²	MAE	MSE	RMSE
Least Median Squares Regression (LMSR)	0.50667	0.04686	0.00482	0.06944
Ridge Regression	0.31491	0.07317	0.01188	0.10897
K-Nearest Neighbors (KNN)	0.60164	0.03127	0.00156	0.03948

Figura 11: Resultado das redes propostas ao conjunto de dados “Otimização da topologia do micromisturador passivo”, considerando os índices estatísticos

4. Conclusão

O projeto “Desenvolvimento de um banco de dados com informações de processos de mistura e reação química em microdispositivos” demonstrou-se como uma prática ferramenta para o escalonamento de informações de microdispositivos. Os dados nesse armazenados, são importantes ferramentas para o fomento de novas pesquisas no ramo; a última parte deste projeto visou principalmente demonstrar sua capacidade em junção com a inteligência artificial. Como uma perspectiva de continuidade, elucidada-se a necessidade aprimorar os métodos de regressão utilizadas, sobretudo desenvolver modelos de redes neurais próprias a cada conjunto. Como opinião do autor, nesta etapa deve-se compreender um modelo distinto para cada processo a ser analisado. Quanto as demais utilidades acerca do uso do banco de dados concebido por esta pesquisa, é necessário que o mesmo seja manuseado e posto em prática por usuários. Como qualquer *software* desenvolvido ou em sua primeira versão, colocá-lo à mercê do dia a dia representa o melhor mecanismo de aprimoramento e consolidação da ferramenta como algo útil à comunidade científica.

5. Referências Bibliográficas

- CHOUDHARY, S.; PATKAR, U.** Databases for artificial intelligence. *International Journal of Computer Science and Mobile Computing*, v. 5, p. 67-70, 2016.
- FOGLER, H. S.** Elements of chemical reaction engineering. Upper Saddle River, NJ, Prentice Hall PTR, 2006.
- FOSTER, Elvis C. (et al.).** Database Systems A Pragmatic Approach, 2^o ed, Apress, 2016.
- KLEPPMANN, M.** Designing data-intensive applications, The Big Ideas Behind Reliable, Scalable, and Maintainable Systems. O Reilly, 2016.
- MOAYEDI, H.; AGHEL, B.; FOONG, L. K.; BUI, D. T.** Feature validity during learning paradigms for predicting biodiesel purity. *Fuel*, v. 262, p. 116498, 2020
- ORIMOTO, Y.; WATANABE, K.; YAMASHITA, K.; UEHARA, M.; NAKAMURA, H.; FURUYA, T.; MAEDA, H.** Application of Artificial Neural Networks to Rapid Data Analysis in Combinatorial Nanoparticle Syntheses. *The Journal of Physical Chemistry*, v. 116, p. 17885-17896, 2012.
- R. Lo,** Application of Microfluidics in Chemical Engineering, *Chem. Eng. Process Tech.* 1: 1001 (2013) 1–3. <http://www.jscimedcentral.com/ChemicalEngineering/Articles/chemicalengineering-1-1002.php>.
- SANTANA, H. S.; TORTOLA, D. S.; REIS, E. M.; SILVA, J. L.; TARANTO, O. P.** Transesterification Reaction of Sunflower Oil and Etanol for Biodiesel Synthesis in Microchannel Reactor: Experimental and Simulation Studies. *Chemical Engineering Journal*, v. 302, p. 752–62, 2016.
- SANTANA, H. S.; TORTOLA, D. S.; SILVA Jr, J. L.; TARANTO, O. P.** Biodiesel synthesis in micromixer with static elements. *Energy Conversion and Management*, v. 141, p. 28-39, 2017a.
- SURYAWANSHI, P. L.; GUMFEKAR, S. P.; BHANVASE, B. A.; SONAWANE, S. H. PIMPLAPURE, M. S.** A review on microreactors: Reactor fabrication, design, and cutting-edge application. *Chemical Engineering Science*, v.189, p. 431-448, 2018.
- WHITESIDES, G. M.** The origins and the future of microfluidics. *Nature*, v. 442, p. 368-373, 2006.