UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ENGENHARIA
DEPARTAMENTO DE ENGENHARIA QUÍMICA
TRABALHO DE DIPLOMAÇÃO EM ENGENHARIA QUÍMICA

Predição da viscosidade de líquidos iônicos através da metodologia QSPR

Autor: Vinícius Grazzioli Moscon

Orientador: Rafael De Pelegrini Soares

Porto Alegre, novembro de 2021

Autor: Vinícius Grazzioli Moscon

Predição da viscosidade de líquidos iônicos através da metodologia QSPR

Trabalho de Conclusão de Curso apresentado à COMGRAD/ENQ da Universidade Federal do Rio Grande do Sul como parte dos requisitos para a obtenção do título de Bacharel em Engenharia Química

Orientador: Rafael De Pelegrini Soares

Banca Examinadora:

Prof.ª Dra. Paula Bettio Staudt, Universidade Federal do Rio Grande do Sul

Doutoranda Anne Caroline Belusso, Universidade Federal do Rio Grande do Sul

Porto Alegre

AGRADECIMENTOS

Dedico este trabalho aos meus pais e meu irmão. Apesar de todas dificuldades encontradas durante a jornada universitária, sempre tive o apoio e motivação necessária para me dedicar, principalmente, ao meu desenvolvimento profissional.

Agradeço aos meus amigos por estarem sempre ao meu lado durante minhas escolhas e me apoiarem ao longo de todos os anos.

Aos professores agradeço por toda a paciência e ensinamentos durante toda a minha vida acadêmica.

RESUMO

Debates a respeito de preservação e sustentabilidade tem aumentado ao longo dos anos, entre as pautas abordadas está o impacto causado por solventes orgânicos ao ambiente. Como alternativa de substituição destes solventes, estão surgindo, cada vez mais em destaque, os líquidos iônicos, já que a grande maioria deles possui pressão de vapor desprezível em temperatura ambiente. Entretanto, o desenvolvimento e estudo das propriedades destes líquidos pode ser caro e demorado. Uma das propriedades fundamentais em qualquer processo químico, que deve ser conhecida, é a viscosidade do fluído. Com esta motivação, neste trabalho foi desenvolvido um modelo de predição de viscosidade de líquidos iônicos baseado na relação quantitativa estrutura-propriedade (QSPR). Esta relação é uma abordagem eficaz para determinar uma relação quantitativa entre a viscosidade e a estrutura iônica para líquidos iônicos, pois vincula matematicamente as propriedades físicas ou químicas com a estrutura de uma molécula. Para desenvolvimento do trabalho, foram utilizados descritores moleculares como variáveis de entrada do modelo matemático gerado pela técnica de aprendizado de máquina. Esses descritores foram calculados, previamente, pelo pacote computacional EnalosMold2, dentro do software KNIME. Foram selecionados os descritores mais correlacionados com a variável de saída, a viscosidade. Através de um aprendizado de máquina, foi gerado um modelo que é capaz de predizer a viscosidade dos líquidos iônicos. Esse modelo foi avaliado por métricas de erro e pelos critérios de Tropsha. O resultado foi bastante satisfatório e todos os critérios de Tropsha foram aceitos, logo, conclui-se que o modelo gerado através da abordagem QSPR é capaz de predizer o fenômeno estudado.

Palavras-chave: *QSPR*, descritores, viscosidade, predição, aprendizado de máquina.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

LIs – Líquidos iônicos;

QSPR - Relação quantitativa estrutura-propriedade;

MSE – Erro quadrático médio;

RMSE - Raiz do erro quadrático médio;

MAE – Erro médio absoluto;

MAPE - Erro percentual absoluto médio.

SUMÁRIO

1 Introdução

A busca por processos químicos mais limpos é uma pauta cada vez mais importante e necessária. Solventes orgânicos são utilizados frequentemente em processos industriais, como por exemplo: tolueno, benzeno, diclorometano, acetonitrila, metanol e etanol. Estes solventes possuem uma característica em comum: se mantém na condição de líquido em uma estreita faixa de temperatura, portanto, são relativamente voláteis nas condições de processo industrial em que são descarregados na atmosfera (ALLEN; SHONNARD, 2001). Essas emissões têm sido associadas a uma série de efeitos negativos, incluindo mudança climática global, má qualidade do ar urbano e doenças humanas.

O profissional da engenharia química deve continuar fornecendo a população os produtos necessários para sustentar um alto padrão de vida e, ao mesmo tempo, reduzir significativamente o impacto ambiental dos processos que utilizamos. Os líquidos iônicos (LIs – em Inglês Ionic Liquids) são uma nova classe de solventes que surgiu nos últimos vinte anos e pode se tornar uma aliada fundamental para enfrentar este desafio (YU et al., 2012). Estes líquidos agem como bons solventes orgânicos, dissolvendo espécies polares e não polares. Em muitos casos, eles têm um desempenho muito melhor do que os solventes comumente usados. Talvez a característica mais intrigante desses compostos seja que, embora sejam líquidos em seu estado puro à temperatura ambiente, grande parte deles não têm pressão de vapor. Assim, boa parte dessas substâncias simplesmente não evaporam em amplas faixas de temperatura, portanto, não geram emissões indesejadas. Muitos desses compostos são líquidos em faixas de temperatura incrivelmente grandes, o que sugere que eles podem ser usados em condições de processamento exclusivas. Eles são sais orgânicos, cujos cátions e ânions podem ser variados virtualmente à vontade para alterar suas propriedades químicas e físicas (BRENNECKE; MAGINN, 2001).

Geralmente, os líquidos iônicos possuem uma viscosidade maior do que os solventes orgânicos convencionais, dando origem a alguns problemas nos processos químicos, variando de um efeito negativo sobre os requisitos de energia até a redução da taxa de transferência de massa e/ou calor na reação e separação. No entanto, LIs altamente viscosos são utilizados em algumas aplicações, como fases estacionárias para cromatografia gás-líquido. Portanto,

preparar líquidos iônicos com valor de viscosidade ideal é certamente uma necessidade e o entendimento da relação entre a viscosidade e a estrutura iônica também é necessário para projetar e sintetizar mais racionalmente um líquido (BRENNECKE; MAGINN, 2001).

A abordagem de relação quantitativa estrutura-propriedade (QSPR – em Inglês Quantitative Structure Property Relationship) pode ser descrita como um método estatístico de análise de dados para desenvolver modelos que possam predizer corretamente determinada propriedade de compostos baseados em sua estrutura química (TROPSHA; GRAMATICA; GOMBAR, 2003). Com as estruturas dos LIs definidas, é possível o cálculo dos descritores moleculares e então, utilizar a abordagem QSPR para determinação de viscosidade de LIs (YU et al., 2012).

O presente trabalho tem como objetivo desenvolver um modelo QSPR que prevê a viscosidade de líquidos iônicos através de suas estruturas moleculares descritas. Para descrever matematicamente as estruturas iônicas será utilizado o pacote computacional *EnalosMold2* que calcula os descritores moleculares de cada íon. Através dos valores que descrevem os íons e um número relativamente grande de dados de viscosidade de LIs, é possível gerar um modelo de regressão que prevê o valor da viscosidade dos LIs.

Para avaliar a robustez e capacidade de predição do modelo gerado, é necessário comparar o resultado obtido através dele com o resultado obtido experimentalmente.

2 Revisão Bibliográfica

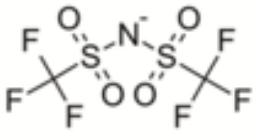
2.1 Líquidos iônicos

Líquidos iônicos são uma nova classe de compostos não moleculares, sendo formados, geralmente, por um cátion orgânico e um ânion inorgânico ou orgânico (BERTOTI; FERREIRA, 2009). Por definição os LIs possuem um ponto de fusão inferior a 100 ºC e grande parte destes apresenta pressão de vapor desprezível em temperatura ambiente, alta estabilidade térmica e propriedades eletrolíticas interessantes (NEBIG; GMEHLING, 2011).

O primeiro líquido iônico sintetizado foi o nitrato de etilamônio, descoberto por Walden, em 1914, que apresenta um ponto de fusão de 12 ºC. A formação de um líquido iônico consiste em uma reação de complexação de espécies aniônicas com compostos neutros. Os líquidos iônicos inicialmente foram desenvolvidos para serem utilizados como eletrólitos para baterias.

A partir da década de 70, líquidos iônicos passaram a ser utilizados como catalisadores em reações orgânicas de alquilação e acilação de olefinas e como solventes para diferentes reações. Chum e Koch, trabalharam com misturas de etilpiridíneo e cloreto de alumínio, formando uma nova classe de cátions para formar LIs (LI CHUM, 1975). Um dos problemas desta classe de LIs, da década de 70 e 80, é a instabilidade em meio aquoso. Para remediar esta situação, na década de 90, foram substituídos os ânions complexos por tetrafluoroborato e hexafluorofosfato, apresentando maior estabilidade química, porém, apresentavam baixa capacidade de coordenação, ainda na década de 90, dois grupos sugeriram que o ânion bis(trifluorometanosulfonil)imideto), apresentado na Figura 1, fosse usado, já que garantia alta hidrofobicidade e elevada estabilidade química, e desde então ele vem sendo o mais utilizado (MARTINS, 2014).

Figura 1. Ânion bis(trifluorometanosulfonil)imideto) (MARTINS, 2014).



Em 1997, Seddon foi o primeiro a sugerir que os líquidos iônicos, devido as suas propriedades, tem um grande potencial na química verde. Eles vêm sendo cada vez mais utilizados em diversos campos do conhecimento, por exemplo: como solventes em catálise bifásica, em eletroquímica, como solventes para extração líquido-líquido, como solventes para reações orgânicas, como fase estacionária para cromatografia gasosa, entre outros.

Comparado aos solventes orgânicos convencionais, o uso de líquidos iônicos tem inúmeras vantagens, determinadas pela combinação original de suas propriedades, são elas: diminuição da perda de solvente em processos químicos, menor agressividade ao meio ambiente e também, maior seletividade em processos de separação (BERTOTI; FERREIRA, 2009). Além disso, uma das grandes vantagens dos líquidos iônicos é a possibilidade de projetar o LI de acordo com a necessidade do meio reacional.

2.2 Descritores moleculares

As propriedades físico-químicas de compostos dependem de suas estruturas moleculares, onde o termo estrutura inclui aspectos topológicos, eletrônicos e geométricos. Para se obter relações entre estruturas químicas e as propriedades, utilizando métodos computacionais, é necessário encontrar representações apropriadas dessas estruturas moleculares (NELSON; SEYBOLD, 2001).

Para descrever a estrutura molecular na forma de um número, foram desenvolvidos muitos termos matemáticos chamados de descritores moleculares, estes descritores refletem propriedades moleculares simples e podem trazer algum tipo de entendimento da característica estrutural intrínseca relacionada com a natureza da propriedade que está sob observação (LIU, 2000).

Um descritor molecular é o resultado obtido através de um processamento lógico e matemático, aplicado às informações químicas codificadas através da representação de uma molécula. Este processamento transforma estas informações codificadas em um valor numérico associado a uma determinada propriedade molecular importante, como por exemplo, viscosidade ou até mesmo ponto de fusão (CONSONNI; TODESCHINI; PAVAN, 2002).

Essa representação é útil, já que as propriedades de uma molécula também são registradas como um número simples.

O mundo de descritores pode ser dividido em dois conjuntos básicos: índices topológicos, e descritores eletrônicos, geométricos e combinados. Índices topológicos são índices derivados exclusivamente da conectividade e composição da estrutura, e teve seu estudo pioneiro em 1947, realizado por Wiener. Descritores derivados puramente da geometria molecular ou distribuição parcial de carga foram designados, respectivamente, como geométricos ou eletrônicos. Descritores que são calculados usando simultaneamente conectividade e estrutura eletrônica ou informações sobre geometria e estrutura eletrônica, foram chamados de descritores combinados (KATRITZKY; GORDEEVA, 1993). Exemplos práticos de descritores: número de carbonos, distância média entre carbonos, densidade de carga, entre outros.

A escolha dos descritores moleculares apropriados para serem aplicados na metodologia QSPR é um dos maiores desafios, pois para se obter uma correlação significativa sob o ponto de vista estatístico, é necessário uma escolha adequada de descritores (ALKER, 2003).

O sucesso da metodologia QSPR depende do cálculo dos descritores moleculares. Um software gratuito que pode ser utilizado para cálculo dos descritores é o *EnalosMold2,* que calcula 777 descritores por molécula, criando uma importante e robusta base de dados de variáveis independentes para cálculo e validação do modelo de predição (AFANTITIS; TSOUMANIS; MELAGRAKI, 2020).

2.3 Metodologia QSPR

Estudos das relações quantitativas entre a estrutura molecular e algum tipo de propriedade físico-química são de grande importância na química moderna. O objetivo principal da metodologia QSPR é simplificar a procura por compostos com propriedades desejadas. Existem duas metodologias para encontrar estes compostos, são elas:

a) Utilização da intuição e experiência química;

b) Utilização de descritores moleculares e modelos de previsão.

Uma vez que a correlação entre estrutura/propriedade é encontrada, um grande número de compostos, incluindo aqueles que ainda não foram sintetizados, podem ser facilmente examinados no computador com o objetivo de selecionar estruturas com as propriedades físico-químicas desejadas. Desta forma, é possível selecionar os compostos mais promissores para síntese e testes em laboratórios, evitando assim, desenvolvimento de compostos desnecessários, economizando tempo e dinheiro. Estudos envolvendo a metodologia QSPR são considerados ótimas ferramentas para acelerar e obter êxito no processo de desenvolvimento de novas moléculas a serem utilizadas (ARROIO; HONÓRIO; SILVA, 2010).

A metodologia QSPR possui diferentes formas de desenvolvimento. Neste presente trabalho será abordada a metodologia citada por Plavsic (1993), que é apresentada na Figura 2 e desenvolvida da seguinte forma:

1) Obtenção dos dados

Nesta etapa é de extrema importância a obtenção de dados confiáveis e precisos, pois a qualidade de todas etapas subsequentes depende dessa fonte de dados;

2) Caracterização da estrutura molecular

São calculados os descritores moleculares podendo ser eletrônicos, geométricos ou topológicos. Estes devem ser selecionados e computados adequadamente;

3) Modelo QSPR

Nesta etapa, o conjunto de dados experimentais e descritores moleculares são correlacionados por meio de métodos matemáticos computacionais resultando numa expressão algébrica aceitável, o modelo deve ser avaliado estatisticamente bem como sua qualidade;

4) Habilidade da previsão do modelo

Os valores das propriedades moleculares de interesse são previstos para moléculas que não fazem parte do grupo de treinamento usando o modelo de QSPR obtido inicialmente.

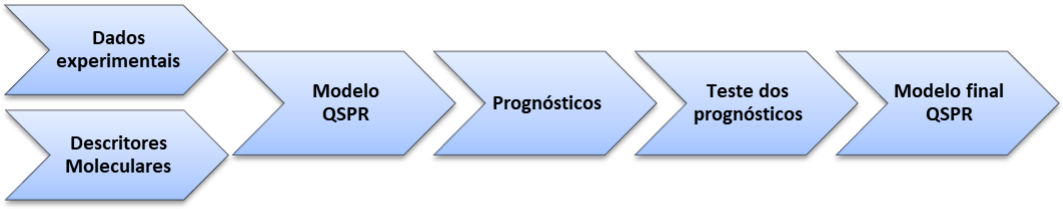
5) Validação do modelo

Com a parcela de dados experimentais que ficaram de fora do grupo de treinamento do modelo, é possível confirmar os prognósticos realizados pelo modelo QSPR. Caso os testes realizados não apresentem um resultado satisfatório, o modelo deve ser revisado e repetido a partir da etapa 3, até a obtenção de um modelo adequado;

6) Modelo final QSPR

Se os testes realizados confirmam os prognósticos, o modelo QSPR é aceito na sua forma final com seus dados estatísticos.

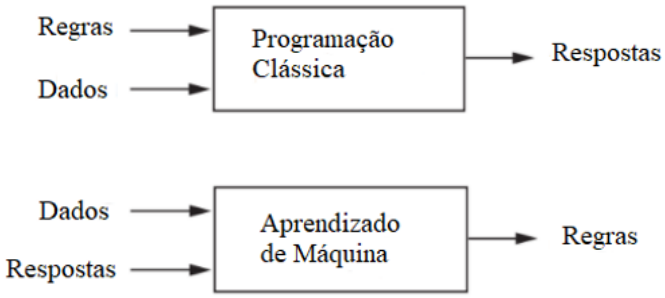
Figura 2. Fluxograma da abordagem QSPR.



2.4 Aprendizado de máquina

Segundo Chollet e Lorenzen (2018), a partir de 1990 tivemos um novo paradigma de programação no desenvolvimento de sistemas de inteligência artificial. A abordagem simbólica, amplamente utilizada no desenvolvimento de sistemas inteligentes, consistia na definição de regras baseadas nos dados disponíveis para geração de respostas. Porém novas técnicas foram desenvolvidas, as regras eram encontradas de forma automatizada a partir de dados e suas respostas. Conforme a Figura 3 a seguir, podemos ver o comparativo entre essas duas abordagens.

Figura 3. Aprendizado de máquina: um novo paradigma de programação (CHOLLET, 2018).



Segundo Géron (2019), aprendizado de máquina é uma área da programação de computadores que estuda técnicas capazes de obter sistemas inteligentes a partir dos dados sem ser explicitamente programado. Para que tal sistema seja desenvolvido são necessários 3 elementos básicos: experiência (representada pelo conjunto de dados), tarefa (atividade a ser realizada pelo algoritmo desenvolvido), e por fim uma medida de desempenho.

O desenvolvimento desse tipo de sistema se baseia na aplicação de dados históricos (experiência) aplicados à execução de uma atividade (tarefa) e para mensurar a eficiência desse sistema utiliza-se de métricas capazes de mostrar o quão bem tal atividade foi executada. Podemos exemplificar da seguinte forma: sendo experiência as características dos imóveis à venda (como número de quartos, localização, tamanho, entre outros atributos) e seu valor de mercado como tarefa, teremos um sistema capaz de precificar o valor de revenda de um imóvel baseado em seus atributos, e por fim como medida de desempenho a comparação entre o valor real de revenda e o valor predito pelo sistema (GÉRON, 2019).

Para resolver o problema citado poderíamos utilizar técnicas de programação tradicional, criando um sistema de regras com base nos padrões encontrados nos dados (ex: a cada banheiro o valor do imóvel aumenta em x unidades). Seriam analisadas todas as características dos imóveis e então montada uma lista de verificações para cada uma delas a fim de compor o sistema, tornando-o um grande conjunto de regras complexas. Conforme citado por Géron (2019), a desvantagem de sistemas assim é nítida quando temos uma variação de comportamento na população dos dados, tornando o conjunto de regras obsoleto (ex: variação da valorização de um bairro faria com que o conjunto de regras tivesse que ser revisto devido a esse fenômeno).

Segundo Grus (2016), o ganho na utilização de técnicas de aprendizado de máquina está na aplicação de algoritmos capazes de identificar de forma automática um conjunto de regras e padrões dentro dos dados para resolver uma atividade y. Tal característica se acentua em problemas mais complexos ou que exigem longas listas de regras.

3 Metodologia e Aspectos computacionais

Neste presente capítulo serão abordados os métodos e programas utilizados para coleta e preparação de dados para utilização do aprendizado de máquina. Além disso, será apresentado o fluxograma utilizado na metodologia QSPR.

3.1 Dados experimentais

Para desenvolvimento da metodologia QSPR, é necessário um banco de dados grande suficiente para gerar um modelo de predição consistente. No presente trabalho, foram coletados dados experimentais de viscosidade de diversos líquidos iônicos apresentados no trabalho de YU (2012). A tabela continha 306 cátions e 138 ânions com suas devidas estruturas químicas, todos presentes em algum dos 5046 líquidos iônicos presentes no conjunto de dados.

Foi realizado um pré-tratamento de dados, onde foram considerados apenas íons em solução de água e, também, foram considerados os diferentes métodos de medição de viscosidade experimental para desenvolver o modelo QSPR. Entretanto, não foram utilizados os dados de composição dos cátions e ânions. Esses fatores podem influenciar no resultado da viscosidade experimental.

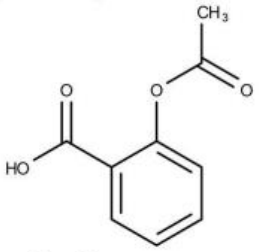
3.2 Descritores moleculares

Para cálculo dos descritores moleculares, é necessário que o pacote computacional utilizado seja capaz de interpretar as estruturas químicas envolvidas no estudo, como apresentado na Figura 4. Para isso, foi utilizado o software Open Babel, que é uma caixa de ferramentas química projetada para falar as várias linguagens dos dados químicos. Este software é um projeto aberto e colaborativo que permite a qualquer pessoa pesquisar, converter, analisar ou armazenar dados de modelagem molecular, química, bioquímica ou áreas relacionadas.

Foram gerados diversos arquivos de cátions e de ânions no formato *.mol* , uma linguagem computacional de estruturas químicas. Para realizar a conversão de estruturas químicas para arquivos .mol, é necessário desenhar todos os íons envolvidos no trabalho no software Open Babel. Para conferência da estrutura química foi utilizado o software Avogrado, que é um programa que ilustra a molécula codificada na linguagem .mol. Para este trabalho, todos os arquivos gerados tinham suas estruturas representadas em 2D devido ao software de cálculo de descritores aceitar apenas esse tipo de formato. Existem softwares pagos que utilizam estruturas representadas em 3D, gerando um melhor desempenho em seus modelos.

Figura 4. Métodos de representação de estruturas químicas, utilizando o ácido acetilsalicílico (Aspirina) como exemplo (ALVES *et al.*, 2017).

A) Representação bidimensional



B) Representação tridimensional

C) Notações lineares

SMILES: CC(=O)Oc1ccccc1C(=O)O

SMARTS: [#6]-[#6](=O)-[#8]-[#6]-1=[#6]-[#6]=[#6]-[#6]=[#6]-1-[#6](-[#8])=O

InChIKey: BSYNRYMUTXBXSQ-UHFFFAOYSA-N

D) Amostra de um arquivo CT (SDfile)

```
Mrv16a2402031713302D

 13 13  0  0  0  0            999 V2000
    1.4289    3.3000    0.0000 C   0  0  0  0  0  0  0  0  0  0
    1.4289    2.4750    0.0000 C   0  0  0  0  0  0  0  0  0  0
    2.1434    2.0625    0.0000 O   0  0  0  0  0  0  0  0  0  0
    0.7145    2.0625    0.0000 O   0  0  0  0  0  0  0  0  0  0
    0.7145    1.2375    0.0000 C   0  0  0  0  0  0  0  0  0  0
    1.4289    0.8250    0.0000 C   0  0  0  0  0  0  0  0  0  0
    1.4289   -0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0
    0.7145   -0.4125    0.0000 C   0  0  0  0  0  0  0  0  0  0
    0.0000    0.0000    0.0000 C   0  0  0  0  0  0  0  0  0  0
    0.0000    0.8250    0.0000 C   0  0  0  0  0  0  0  0  0  0
   -0.7145    1.2375    0.0000 C   0  0  0  0  0  0  0  0  0  0
   -0.7145    2.0625    0.0000 O   0  0  0  0  0  0  0  0  0  0
   -1.4289    0.8250    0.0000 O   0  0  0  0  0  0  0  0  0  0
  1  2  1  0  0  0  0
  2  3  2  0  0  0  0
  2  4  1  0  0  0  0
  4  5  1  0  0  0  0
            ...
```

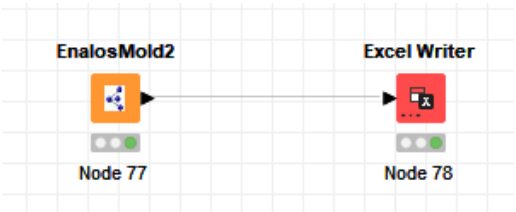
O conjunto de dados continha líquidos iônicos com até 3 íons envolvidos e em solução de diversos solventes, e lembrando que, para fins de simplificação, foram utilizados apenas líquidos iônicos com 2 íons envolvidos e em mesma solução (água). Com essas condições chegamos ao total de 100 ânions, 296 cátions e 2715 líquidos iônicos.

3.3 Fluxograma para geração do modelo QSPR

Para trabalhar com os dados de viscosidade experimental e arquivos de cátions e ânions foi utilizada a plataforma KNIME Analytics, que é um software de código aberto de análise de dados, construção de relatórios e integração de dados. O KNIME integra vários componentes para aprendizado de máquina e mineração de dados por meio de seu conceito de *pipelining* modular. Ele contém todas as principais técnicas de data *wrangling* e aprendizado de máquina baseados em programação visual.

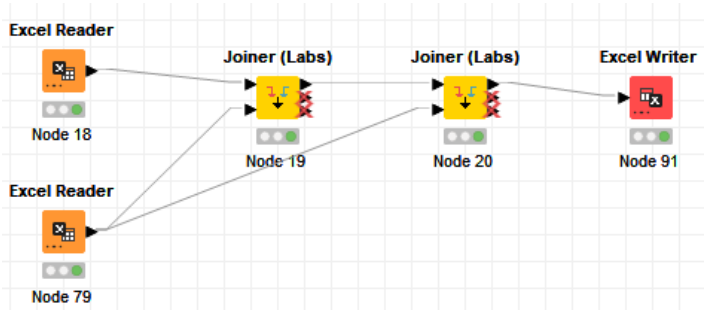
A primeira etapa foi realizar a leitura dos arquivos de cátions e ânions através do módulo *EnalosMold2*, do pacote Mold2, conforme Figura 5. Esse pacote é capaz de calcular um conjunto grande e diversificado de descritores moleculares que codificam informações de estrutura química bidimensional. A análise comparativa dos descritores Mold2 com aqueles calculados a partir de softwares comerciais em vários conjuntos de dados publicados, demonstrou que, os descritores Mold2 transmitem informações estruturais suficientes. Além disso, modelos melhores foram gerados usando os descritores Mold2 do que os gerados por pacotes de softwares comerciais (MELAGRAKI; AFANTITIS, 2013). Este software público é desenvolvido pelo Centro de Bioinformática, que é liderado pelo Dr. Weida Tong, no *National Center for Toxicological Research* (NCTR). Vale ressaltar que, apesar deste software calcular os descritores, ele não indica publicamente o significado de cada um deles.

Figura 5. Fluxograma para cálculo dos descritores no software KNIME.



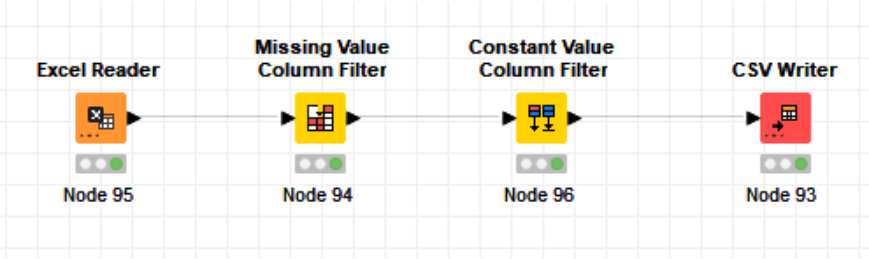
Foram, então, calculados 777 descritores para cada cátion e para cada ânion. Após esta etapa, foram unidos cátions e ânions, conforme Figura 6, formando o líquido iônico que possuía resultado de viscosidade experimental, gerando uma tabela de 2715 líquidos iônicos com 1556 descritores (foram adicionados dados de temperatura e pressão). Os descritores são considerados variáveis de entrada enquanto o logaritmo natural da viscosidade experimental é considerado a variável de saída.

Figura 6. Fluxograma de geração de conjunto de dados no software KNIME.



É de extrema importância que esses descritores sejam avaliados pois alguns não são representativos para a variável de saída, ou seja, então nenhuma informação sobre a variável de saída pode ser obtida conhecendo determinada variável de entrada. Outro ponto é que não é interessante ter um grande número de variáveis para cálculo da equação, já que o modelo tem um risco crescente de *overfitting* com o aumento do número de colunas. Primeiramente, foi necessário eliminar todas as colunas que continham um valor constante para todos líquidos iônicos ou que por algum motivo não foi possível calcular o valor dos descritores. Foram retiradas 470 colunas na operação apresentada na Figura 7.

Figura 7. Fluxograma para avaliação dos dados no software KNIME.

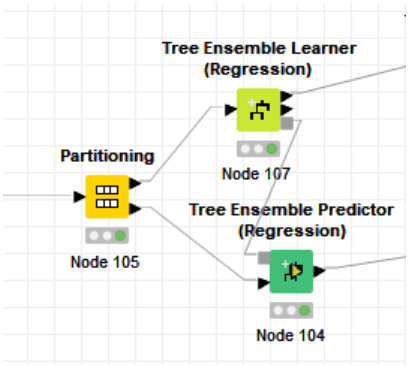


Ainda na avaliação dos descritores foi realizado uma *feature selection* do *data set.* Para isso foi utilizado software WEKA (FRANK *et al.*, 2004). O software utiliza alguns testes estatísticos para realizar essa seleção, porém aceita dados apenas no formato .csv. Como o KNIME é um software montado em módulos de códigos, ele possui um módulo em que é possível fazer essa transformação e, assim então, abrir o arquivo gerado pelo software WEKA.

Para a seleção de variáveis, no software WEKA, foi selecionado a opção de em que o método de seleção era o *BestFirst* com parâmetros WEKA padrão, baseado nos autores BARBOSA e STEFANI (2013). Após a conclusão deste procedimento, restaram apenas 18 descritores representativos para geração do modelo.

Com o *dataset* preparado para geração do modelo, é necessário realizar uma divisão em dois grupos de dados (70%/30%), um grupo será utilizado para o aprendizado da máquina e o outro será utilizado para avaliação do modelo gerado, conforme apresentado na Figura 8.

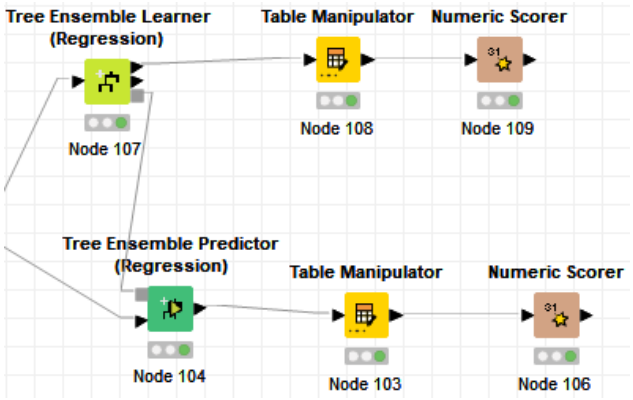
Figura 8. Fluxograma para geração de modelo matemático através da técnica de aprendizado de máquina no software KNIME.



Para desenvolvimento do modelo de regressão, pode ser escolhido a metodologia mais eficiente, o que depende, então, dos resultados gerados e avaliados. Neste trabalho foi selecionado o *Tree Ensemble Learner*, que é um modelo preditivo, um tipo de árvore de decisão. Esse tipo de modelo usa soma de quadrados e análise de regressão para prever valores do campo de destino. As previsões são baseadas em combinações de valores nos campos de entrada.

Com o modelo de regressão concluído, é feita a avaliação do resultado alcançado, conforme Figura 9, a qual será apresentada nos tópicos seguintes.

Figura 9. Fluxograma de avaliação do modelo gerado no software KNIME.



3.4 Métrica R² - Coeficiente de determinação

A medida do coeficiente de determinação é uma métrica que visa expressar a quantidade da variância dos dados que é explicada pelo modelo construído. Em outras palavras, essa medida calcula qual a porcentagem da variância que pôde ser prevista pelo modelo de regressão e, portanto, nos diz o quão próximo os valores previstos pelo modelo estão dos valores reais (medidos em experimentos). O coeficiente de determinação é definido da seguinte forma:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \qquad (1)$$

Onde na equação (1) y é o valor experimental, $\hat{y}$ é o valor previsto e $\bar{y}$ é o valor médio dos dados experimentais.

Uma das desvantagens do R² é que, por definição, o coeficiente de determinação é enviesado. Isso ocorre uma vez que os otimizadores dos algoritmos de regressão utilizam da correlação dos dados de forma a incrementar o valor de R², o que causa um aumento sistemático desse valor conforme novas medidas são adicionadas.

Para contornar este problema, também foi avaliado o R^2-ajustado, que partindo do mesmo princípio do R^2, busca representar a porcentagem da variância que pode ser contemplada pelo modelo de regressão. Entretanto, esse valor não demonstra um viés devido ao acréscimo de dados ou *features* no modelo.

3.5 Métricas dos erros (MSE, RMSE, MAE, MAPE)

Uma métrica muito utilizada para avaliar o modelo de regressão é o erro quadrático médio (MSE). Basicamente, essa métrica calcula a diferença entre o valor predito e o valor real, eleva ao quadrado, faz-se a mesma coisa com todos os outros pontos, soma-os e divide-se pelo número total de valores preditos. Logo, quanto maior o valor de MSE, pior o modelo.

Uma das desvantagens do MSE é que, para a predição de valores de unidade u, a unidade do MSE seria u^2. Para contornar esse problema, existe outra métrica utilizada, a raiz do erro quadrático médio (RMSE). O RMSE é calculado através da raiz quadrada do MSE, acertando a unidade do erro. Da mesma forma que o MSE, quanto maior o valor de RMSE, pior é o modelo.

É possível mensurar a performance do modelo através do erro médio absoluto (MAE), que consiste em calcular a média absoluta da variação entre valor predito e valor medido. Um modelo perfeito teria um resultado de MAE de zero, logo, quanto menor o valor de MAE, melhor será a robustez do modelo.

Em contraste com as métricas anteriores, existe o erro percentual absoluto médio (MAPE), que é uma medida bastante intuitiva tanto para a interpretação do programador, quanto para a comunicação de resultados com pessoas sem conhecimento técnico. Por exemplo, ter um MAPE = 30% significa que, em média, nosso modelo faz previsões que erram por 30% do valor real.

Os valores de MSE, RMSE, MAE e MAPE são definidos da seguinte forma:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{(n)}} \quad (2)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (3)$$

$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i} \quad (4)$$

$$MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{(n)} \quad (5)$$

Onde nas equações (2), (3), (4) e (5) y é o valor experimental, $\hat{y}$ é o valor previsto e n é o número de observações.

3.6 Validação cruzada (K-fold)

Em machine learnig é muito comum dividirmos o conjunto de dados para treino do modelo e para teste. Porém, essa divisão ocorre de forma aleatória, e cada vez que o modelo é gerado, resulta em um resultado diferente. Para resolver este problema, é possível utilizarmos a metodologia de validação cruzada K-fold.

K-fold é o método de validação cruzada mais conhecido e utilizado. O método consiste em dividir o conjunto de dados em k partes, usando k-1 partes para treino e a parte remanescente para teste, fazendo isso k vezes. Em cada uma das k vezes, testa-se o modelo com um fold diferente calculando a métrica escolhida para avaliação do modelo.

3.7 Validação do modelo

A abordagem QSPR é, basicamente, um modelo matemático gerado por técnicas de aprendizado de máquina baseado em um conjunto de dados. Essa abordagem deve ser avaliada estatisticamente para definir a robustez do modelo gerado e, também, sua capacidade de predição.

O primeiro indicador de sucesso de um modelo QSPR é medir a qualidade de ajuste nos dados de treinamento disponíveis. Os critérios mais comumente utilizados são o coeficiente de determinação e a métrica dos erros.

Para avaliar o modelo gerado Tropsha, Gramatica e Gombar (2003) citam que é considerado um modelo preditivo o que seguir as seguintes condições:

$$R^2 > 0{,}6 \quad (6)$$

$$Rcvext^2 > 0{,}5 \quad (7)$$

$$\frac{(R^2 - {R_0}^2)}{R^2} < 0{,}1 \quad (8)$$

$$\frac{(R^2 - {R_0}'^2)}{R^2} < 0{,}1 \quad (9)$$

$$abs\left(R - {R_0}'^2\right) < 0{,}1 \quad (10)$$

$$0{,}85 < k < 1{,}15 \quad (11)$$

$$0{,}85 < k' < 1{,}15 \quad (12)$$

Onde R_0 e ${R_0}'$ são calculados da seguinte forma:

$${R_0}^2 = 1 - \sum_{n=1}^{n} \frac{(\hat{y}_i - {\mathrm{y}_i}^{r_0})^2}{(\hat{y}_i - \bar{\hat{y}})} \quad (13)$$

$${R_0}'^2 = 1 - \sum_{n=1}^{n} \frac{(\mathrm{y}_i - {\hat{y}_i}^{r_0})^2}{(\mathrm{y}_i - \bar{\mathrm{y}})} \quad (14)$$

O valor de ${\mathrm{y}_i}^{r_0}$ e ${\hat{y}_i}^{r_0}$ são definidos como:

$${\mathrm{y}_i}^{r_0} = k\,\hat{\mathrm{y}}_i \quad (15)$$

$${\hat{y}_i}^{r_0} = k'\ \mathrm{y}_i \quad (16)$$

Os valores de *k* e *k'* são definidos como os coeficientes angulares das regressões lineares que passam pela origem (dos valores previstos x observados e valores observados x previstos, respectivamente).

Para que o modelo tenha alta capacidade de predição, é necessário que os valores de R_0 e ${R_0}'$ tenham valores similares.

4 Resultados e Discussões

Neste capítulo serão apresentados os resultados obtidos através do modelo de regressão e de toda a metodologia QSPR. Um modelo QSPR é uma representação teórica, logo, para avaliar sua robustez, ou seja, sua capacidade de predição, é necessário seguir alguns critérios estatísticos.

No presente trabalho foram avaliados os seguintes critérios estatísticos: coeficiente de determinação entre os valores experimentais e previstos pelo modelo (R^2), avaliação das métricas dos erros, validação cruzada (k-fold), validação com um conjunto de dados externo e, também, avaliados os critérios definidos por Tropsha, Gramatica e Gombar (2003).

4.1 Avaliação da métrica R^2 - Coeficiente de determinação

Os valores de R^2 e R^2-ajustado foram calculados através de um código na linguagem Python, que é uma linguagem de programação de alto nível, dinâmica e orientada a objetos. O código foi desenvolvido para que os valores de R^2 e R^2-ajustado sejam calculados para os dados de treinamento e, também, dados de teste, conforme Figura 10.

Figura 10. Código no Python para cálculo de métrica R^2.

Calculo coeficiente de determinação R²

```
#Tamanho do conjunto de dados
n_train = y_train.shape[0]
n_test = y_test.shape[0]

#Quantidade de variáveis preditoras
p = 10

#Calculando r2 do treino
r2_train = r2_score(y_train, y_train_pred)

#Calculando r2 do teste
r2_test = r2_score(y_test, y_test_pred)

#Calculando r2 ajustado do treino
r2_train_adj = 1-(1-r2_train)*(n_train-1)/(n_train-p-1)

#Calculando r2 ajustado do teste
r2_test_adj = 1-(1-r2_test)*(n_test-1)/(n_test-p-1)
```

```
print('R² Treino:', r2_train)
print('R² Ajsutado Treino:', r2_train_adj)
print('-'*50)
print('R² Teste:', r2_test)
print('R² Ajsutado Teste:',r2_test_adj)
```

```
R² Treino: 0.7541403688006324
R² Ajsutado Treino: 0.75283883554918
--------------------------------------------------
R² Teste: 0.8361402474435807
R² Ajsutado Teste: 0.8341021908197447
```

Segundo A. Tropsha Tropsha, Gramatica; Gombar (2003) um valor de $R^2 > 0,6$ é um dos critérios para o modelo ser considerado capaz de predizer o fenômeno estudado.

Existe um forte indício de que a equação criada pelo modelo tem uma boa capacidade de explicar a viscosidade utilizando as variáveis selecionadas. Isso pode ser explicado pelo fato de que o valor de R^2 do teste (R^2 teste = 0,836) é maior do que o valor de R^2 do treinamento (R^2 treinamento = 0,754).

4.2 Avaliação das métricas dos erros (MSE, RMSE, MAE, MAPE)

No presente trabalho, foram calculados, através de um código em linguagem Python, os valores de MSE, RMSE, MAE e MAPE, para o conjunto de dados de treinamento e também, para o conjunto de dados de teste conforme Figura 11:

Figura 11. Código de Python para cálculo das métricas dos erros.

```
#Calculando erro médio quadrado treino
MSE_train = mean_squared_error(y_train, y_train_pred)
print('Treino:',MSE_train)

#Calculando erro médio quadrado treino
MSE_test = mean_squared_error(y_test, y_test_pred)
print('Teste:',MSE_test)
```

```
Treino: 0.5229516756470267
Teste: 0.3272789413545678
```

```
RMSE_train = mean_squared_error(y_train, y_train_pred,squared=False)
print('Treino:',RMSE_train)

RMSE_test = mean_squared_error(y_test, y_test_pred,squared=False)
print('Teste:',RMSE_test)
```

```
Treino: 0.7231539778270093
Teste: 0.5720829846749227
```

```
#Calculando erro médio absoluto treino
MAE_train = mean_absolute_error(y_train, y_train_pred)
print('Treino:', MAE_train)

#Calculando erro médio absoluto teste
MAE_test = mean_absolute_error(y_test, y_test_pred)
print('Teste:',MAE_test)
```

```
Treino: 0.44850982711997656
Teste: 0.37406137955272045
```

```
#Calculando MAPE treino
MAPE_train = np.mean(np.abs((y_train - y_train_pred) / y_train)) * 100
print('Treino:',MAPE_train)

#Calculando MAPE teste
MAPE_test = np.mean(np.abs((y_test - y_test_pred) / y_test)) * 100
print('Teste:',MAPE_test)
```

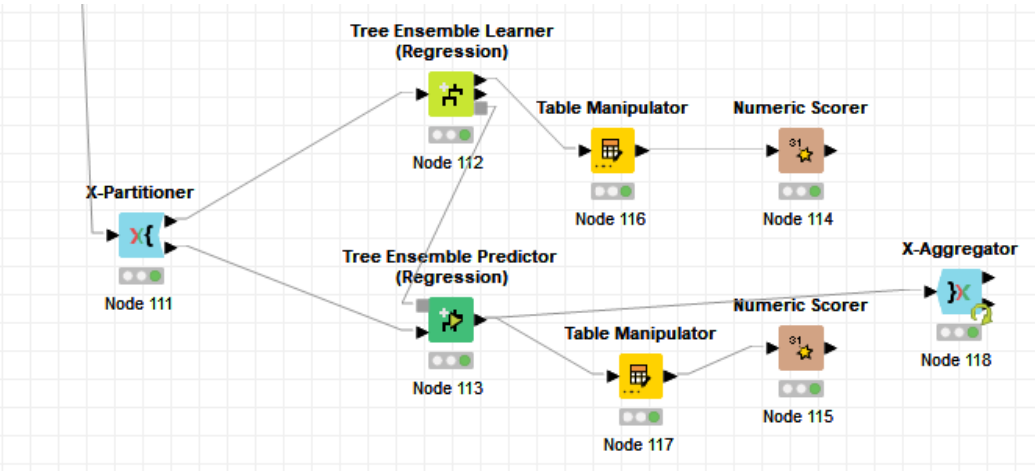
```
Treino: 10.527645596120893
Teste: 8.807779989161709
```

Os valores das métricas de erro para o conjunto de dados de teste foram todos menores do que as métricas para o conjunto de dados de treinamento. Esse fato indica que o modelo gerado consegue predizer o valor da viscosidade dos líquidos iônicos.

4.3 Avaliação da validação cruzada (K-fold)

Para realizar a validação cruzada, K-fold, foi selecionado o valor de 10 para k, um valor comumente encontrado na literatura. Dentro do software KNIME é possível criar um fluxograma que realiza a validação cruzada, apresentado na Figura 12.

Figura 12. Fluxograma de metodologia k-fold no software KNIME.



O valor médio de R^2 para a validação cruzada foi de 0,7879 e os valores de MSE para cada conjunto de testes é apresentado na tabela 1:

Tabela 1. Valores de R^2 e MSE para cada k-fold.

k-fold	MSE	R^2
fold 1	0,35179	0,786
fold 2	0,69823	0,806
fold 3	0,48885	0,783
fold 4	0,35822	0,79
fold 5	0,43397	0,782
fold 6	0,30011	0,778
fold 7	0,33836	0,783
fold 8	0,40994	0,792
fold 9	0,59651	0,791
fold 10	0,37138	0,788

Os valores de R^2 e MSE indicam que independente do conjunto de dados selecionado para teste e para treinamento, o modelo tem um desempenho equivalente e satisfatório.

4.4 Validação do modelo

Para Tropsha, um modelo gerado pela abordagem QSPR tem um desempenho satisfatório e robusto quando os critérios da tabela 2 são seguidos.

Tabela 2. Avaliação dos critérios de Tropsha.

Critério	Avaliação	Resultado
$R^2 > 0{,}6$	OK	$R^2 = 0{,}754$
$Rcvext^2 > 0{,}5$	OK	$Rcvext^2 = 0{,}7879$
$\frac{(R^2 - {R_0}^2)}{R^2} < 0{,}1$	OK	$\frac{(R^2 - {R_0}^2)}{R^2} = -0{,}293$
$\frac{(R^2 - {R_0}'^2)}{R^2} < 0{,}1$	OK	$\frac{(R^2 - {R_0}'^2)}{R^2} = -0{,}311$
$abs(R - {R_0}'^2) < 0{,}1$	OK	$abs(R - {R_0}'^2) = 0{,}014$
$0{,}85 < k < 1{,}15$	OK	$k = 0{,}974$
$0{,}85 < k' < 1{,}15$	OK	$k' = 1{,}009$

Todos os critérios de Tropsha foram aceitos, indicando que através dos descritores moleculares selecionados, é possível gerar um modelo pela metodologia que QSPR é capaz de predizer a viscosidade de líquidos iônicos de maneira satisfatória.

4.5 Apresentação de resultados

Neste tópico serão apresentados alguns exemplos de viscosidade de LIs, obtidos através do modelo gerado pela abordagem QSPR. Para fins de comparação, também serão apresentados os obtidos experimentalmente. Além disso será ilustrado o erro absoluto e desvio relativo de cada ponto, com o objetivo de demonstração do desempenho da abordagem.

Tabela 3. Exemplos de valores de viscosidade gerados pelo modelo, obtidos experimentalmente e seus respectivos erros.

Viscosidade experimental ln[cP]	Viscosidade calculada pelo modelo ln[cP]	Erro Absoluto	Desvio relativo
5,372	5,346	0,027	0,50%
5,215	5,184	0,030	0,58%
5,753	5,937	0,184	3,20%
6,668	6,483	0,186	2,78%
4,197	4,011	0,186	4,44%
2,728	2,541	0,187	6,84%
5,441	5,628	0,188	3,45%
4,404	4,536	0,131	2,98%
3,178	3,310	0,132	4,15%
3,728	3,861	0,133	3,56%
5,026	5,159	0,133	2,64%
4,604	4,471	0,133	2,89%
1,974	1,841	0,133	6,76%
4,357	4,223	0,134	3,07%
3,178	3,044	0,134	4,21%
2,262	2,128	0,134	5,93%
5,945	5,810	0,136	2,28%
1,705	1,841	0,136	7,97%
5,380	5,242	0,138	2,56%
6,677	6,537	0,140	2,10%
4,492	4,351	0,141	3,13%
5,323	5,464	0,141	2,66%
5,176	5,034	0,142	2,75%
4,342	4,485	0,142	3,28%
4,762	4,905	0,143	3,00%
3,653	3,796	0,143	3,91%
4,234	4,091	0,143	3,39%
4,025	4,169	0,144	3,57%
2,766	2,911	0,145	5,23%
5,255	5,110	0,145	2,77%
2,833	2,979	0,146	5,14%
3,938	4,085	0,147	3,74%
4,131	4,279	0,148	3,58%
3,135	3,284	0,149	4,74%
5,475	5,627	0,152	2,77%
4,857	4,785	0,072	1,49%
6,150	5,790	0,360	5,86%
4,984	4,621	0,362	7,27%
3,190	3,553	0,363	11,38%
3,109	3,182	0,073	2,35%
5,030	4,998	0,033	0,65%
4,408	4,208	0,200	4,53%
3,638	3,839	0,201	5,53%
3,861	3,829	0,031	0,81%
4,025	3,993	0,032	0,80%

O desvio relativo entre valor previsto e valor experimental tem um resultado médio de 9%. Além disso, não foi evidenciado nenhum erro sistemático causado por algum íon específico. Esses fatores em conjunto com as outras métricas avaliadas indicam que através do modelo de regressão gerado pela abordagem QSPR é possível predizer viscosidade de LIs apenas com a sua estrutura química.

5 Conclusões e Trabalhos Futuros

5.1 Conclusões

Neste trabalho, foi gerado um modelo de predição de viscosidade de líquidos iônicos através de técnicas de aprendizado de máquina que usa como variáveis de entrada os descritores moleculares de cada composto. Esses descritores moleculares foram calculados por um módulo gratuito do software KNIME, *EnalosMold2*, do pacote Mold2, que requer apenas informações relacionadas à estrutura 2D de cada íon. As variáveis selecionadas foram previamente correlacionadas com a variável de saída através do software WEKA.

O modelo foi avaliado por métricas de erro e, também, por um conjunto de dados externo ao conjunto de dados de treinamento. Por todos os critérios de Tropsha terem sido aceitos, e também, as métricas de erro indicarem que foi gerado um modelo eficiente, conclui-se que as variáveis de entrada foram bem selecionadas e que o modelo gerado é capaz de predizer com robustez o valor de viscosidade de líquidos iônicos.

O método proposto, requer informações relacionadas apenas à estrutura 2D de um composto, podendo ser um auxílio útil para experimentos onerosos e demorados para determinar a viscosidade, já que, é possível predizer a viscosidade de um líquido iônico antes mesmo de seu desenvolvimento, que por muitas vezes, é bastante caro.

5.2 Trabalhos futuros

Por fim, para trabalhos futuros, seria interessante avaliar o que cada um dos descritores moleculares selecionados representa, relacionando assim, como o descritor influencia na propriedade em questão. Além disso, calcular descritores moleculares a partir de estruturas representadas em 3D, geraria um conjunto de dados que, provavelmente, produziria um modelo com maior assertividade. Um outro tipo de técnica que também poderia ser aplicada, é a técnica de aprendizado de máquina profundo (Deep Learning), pois esse tipo de abordagem é capaz de gerar modelos com maior capacidade de aprendizado.

REFERÊNCIAS

AFANTITIS, Antreas; TSOUMANIS, Andreas; MELAGRAKI, Georgia. Enalos Suite of Tools: Enhancing Cheminformatics and Nanoinfor matics through KNIME -. [*s. l.*], n. 2, p. 6523–6535, 2020. Disponível em: https://doi.org/10.2174/0929867327666200727114410

ALKER, J O H N D W *et al.* GUIDELINES FOR DEVELOPING AND USING QUANTITATIVE STRUCTURE – ACTIVITY RELATIONSHIPS. [*s. l.*], v. 22, n. 8, p. 1653–1665, 2003.

ALLEN, David T.; SHONNARD, David R. **Green engineering: Environmentally conscious design of chemical processes and products**. [*S. l.: s. n.*], 2001. Disponível em: https://doi.org/10.1002/aic.690470902

ALVES, Vinicius *et al.* QUIMIOINFORMÁTICA: UMA INTRODUÇÃO. **Química Nova**, [*s. l.*], v. 27, n. 4, p. 631–639, 2017. Disponível em: https://doi.org/10.21577/0100-4042.20170145

ARROIO, Agnaldo; HONÓRIO, Káthia M.; SILVA, Albérico B. F. da. Propriedades químico-quânticas empregadas em estudos das relações estrutura-atividade. **Química Nova**, [*s. l.*], v. 33, n. 3, p. 694–699, 2010. Disponível em: https://doi.org/10.1590/S0100-40422010000300037. Acesso em: 11 maio 2021.

BARBOSA, Rogério; STEFANI, Ricardo. QSPR based on support vector machines to predict the glass transition temperature of compounds used in manufacturing OLEDs. **http://dx.doi.org/10.1080/08927022.2012.717282**, [*s. l.*], v. 39, n. 3, p. 234–244, 2013. Disponível em: https://doi.org/10.1080/08927022.2012.717282. Acesso em: 10 out. 2021.

BERTOTI, Ada Ruth; CARLOS, José; FERREIRA, Netto -. LÍQUIDO IÔNICO [bmim.PF 6] COMO SOLVENTE: UM MEIO CONVENIENTE PARA ESTUDOS POR FOTÓLISE POR PULSO DE LASER. *In*: QUIM. NOVA. [*S. l.: s. n.*], 2009. v. 32.

CHOLLET, François; LORENZEN, Knut. Deep Learning mit Python und Keras : Das Praxis-Handbuch vom Entwickler der Keras-Bibliothek. [*s. l.*], 2018.

CONSONNI, Viviana; TODESCHINI, Roberto; PAVAN, Manuela. Structure/Response Correlations and Similarity/Diversity Analysis by GETAWAY Descriptors. 1. Theory of the Novel 3D Molecular Descriptors. **Journal of Chemical Information and Computer Sciences**, [*s. l.*], v. 42, n. 3, p. 682–692, 2002. Disponível em: https://doi.org/10.1021/ci015504a. Acesso em: 11 maio 2021.

FRANK, Eibe *et al.* Data mining in bioinformatics using Weka. **BIOINFORMATICS APPLICATIONS NOTE**, [*s. l.*], v. 20, n. 15, p. 2479–2481, 2004. Disponível em: https://doi.org/10.1093/bioinformatics/bth261

GÉRON, Aurélien. **Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & TensorFlow**

- Aurélien Géron - Google Livros. [*S. l.*], 2019. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=Z0mvDwAAQBAJ&oi=fnd&pg=PP1&dq=Géron,+Aurélien.+(2019),+Mãos+à+Obra:+Aprendizado+de+Máquina+com+Scikit-Learn+%26+TensorFlow,+Alta+Books,+1º+Edição.&ots=B1uEow6lAX&sig=EZh_JWisctqkDM0H4ZhqbbHsIFI#v=onepage&q. Acesso em: 5 out. 2021.

GRUS, Joel. **Data Science do Zero - Joel Grus - Google Livros**. [*S. l.*], 2016. Disponível em: https://books.google.com.br/books?hl=pt-BR&lr=&id=2LZwDwAAQBAJ&oi=fnd&pg=PT14&dq=Grus,+Joel.+(2016),+Data+Science+do+Zero,+Alta+Books,+1º+Edição.&ots=sqrPLiffFt&sig=S7q52Lj9fUB4pjdMICHKC6zWGRU#v=onepage&q=Grus%2C Joel. (2016)%2C Data Science do Zero%2C Alta Books%2C 1º Edição.&f=false. Acesso em: 5 out. 2021.

HOLBREY, J. D.; SEDDON, K. R. Ionic Liquids. **Clean Technologies and Environmental Policy**, [*s. l.*], v. 1, n. 4, p. 223–236, 1999. Disponível em: https://doi.org/10.1007/s100980050036

KATRITZKY, Alan R; GORDEEVA, Ekaterina V. **Traditional Topological Indices vs Electronic, Geometrical, and Combined Molecular Descriptors in QSAR/QSPR ResearchJ. Chem. Inf. Comput. Sci**. [*S. l.: s. n.*], 1993. Disponível em: https://pubs.acs.org/sharingguidelines. Acesso em: 19 maio 2021.

LI CHUM, Helena *et al.* An Electrochemical Scrutiny of Organometallic Iron Complexes and Hexamethylbenzene in a Room Temperature Molten Salt. **Journal of the American Chemical Society**, [*s. l.*], v. 97, n. 11, p. 3264–3265, 1975. Disponível em: https://doi.org/10.1021/ja00844a081

LIU, Shushen *et al.* Molecular Electronegative Distance Vector (MEDV) Related to 15 Properties of Alkanes. [*s. l.*], 2000. Disponível em: https://doi.org/10.1021/ci0003247. Acesso em: 19 maio 2021.

MARTINS, Vitor Leite. **Líquidos iônicos como eletrólitos para baterias: comportamento eletroquímico de metais e propriedades físico-químicas dos líquidos**. 2014. - Universidade de São Paulo, São Paulo, 2014. Disponível em: https://doi.org/10.11606/T.46.2014.tde-31032014-083507. Acesso em: 10 maio 2021.

NEBIG, Silke; GMEHLING, Jürgen. Prediction of phase equilibria and excess properties for systems with ionic liquids using modified UNIFAC: Typical results and present status of the modified UNIFAC matrix for ionic liquids. **Fluid Phase Equilibria**, [*s. l.*], v. 302, n. 1–2, p. 220–225, 2011. Disponível em: https://doi.org/10.1016/j.fluid.2010.09.021

NELSON, Steven D.; SEYBOLD, Paul G. Molecular structure-property relationships for alkenes. **Journal of Molecular Graphics and Modelling**, [*s. l.*], v. 20, n. 1, p. 36–53, 2001. Disponível em: https://doi.org/10.1016/S1093-3263(01)00099-7

PLAVŠIĆ, Dejan *et al.* On the Harary index for the characterization of chemical graphs. **Journal of Mathematical Chemistry**, [*s. l.*], v. 12, n. 1, p. 235–250, 1993. Disponível em: https://doi.org/10.1007/BF01164638. Acesso em: 25 maio 2021.

RAJAPPAN, Remya *et al.* Quantitative structure-property relationship (QSPR) prediction of liquid viscosities of pure organic compounds employing random forest regression. **Industrial and Engineering Chemistry Research**, [*s. l.*], v. 48, n. 21, p. 9708–9712, 2009. Disponível em: https://doi.org/10.1021/ie8018406. Acesso em: 7 abr. 2021.

TROPSHA, Alexander; GRAMATICA, Paola; GOMBAR, Vijay?K. The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. **QSAR & Combinatorial Science**, [*s. l.*], v. 22, n. 1, p. 69–77, 2003. Disponível em: https://doi.org/10.1002/qsar.200390007

WALDEN, P. **Math-Net.Ru All Russian mathematical portal**. [*S. l.: s. n.*], [*s. d.*]. Disponível em: http://www.mathnet.ru/eng/agreement. Acesso em: 10 maio 2021.

YU, Guangren *et al.* Viscosity of ionic liquids: Database, observation, and quantitative structure-property relationship analysis. **AIChE Journal**, [*s. l.*], v. 58, n. 9, p. 2885–2899, 2012. Disponível em: https://doi.org/10.1002/AIC.12786. Acesso em: 6 out. 2021.

BRENNECKE, Joan F.; MAGINN, Edward J. **Ionic liquids: Innovative fluids for chemical processing.** [S. l.: s. n.], 2001. Disponível em: https://doi.org/10.1002/aic.690471102