

Predictive Modeling of Surface Tension in Chemical Compounds: Uncovering Crucial Features with Machine Learning

Paula J. F. Cala,^a Guilherme G. Dariani,^b Eduardo T. A. Veiga,^c Pedro H. D. Macedo,^c
Amauri J. Paula^{*,a,b} and James M. Almeida^{*,b}

^aDepartamento de Física, Universidade Federal do Ceará (UFC), 60455-900 Fortaleza-CE, Brazil

^bIlum Escola de Ciência, Centro Nacional de Pesquisa em Energia e Materiais (CNPEM),
13087-548 Campinas-SP, Brazil

^cCentro de Ciências Naturais e Humanas, Universidade Federal do ABC, 09280-560 Santo André-SP, Brazil

Surface tension (SFT) can shape the behavior of liquids in industrial chemical processes, influencing variables such as flow rate and separation efficiency. This property is commonly measured with experimental approaches such as Du Noüy ring and Wilhelmy plate methods. Here, we present machine learning (ML) methodologies that can predict the SFT of hydrocarbons. A comparative analysis encompassing k-nearest neighbors, random forest, and XGBoost (extreme gradient boosting) methods was done. Results from our study reveal that XGBoost is the most accurate in predicting hydrocarbon SFT, with a mean squared error (MSE) of 4.65 mN² m⁻² and a coefficient of determination (R²) score of 0.89. The feature importance was evaluated with the permutation feature importance method and Shapley analysis. Enthalpy of vaporization, density, molecular weight and hydrogen content are key factors in accurately predicting SFT. The successful integration of these methodologies holds the potential to impact efficiency in different industry processes.

Keywords: machine learning, physicochemical properties, surface tension, organic molecules, regression

Introduction

Surface tension (SFT) is a key property in fluid mechanics, largely impacting petrochemical operations¹⁻³ and technological applications including microfluidics,^{4,5} cleaning,⁶⁻⁹ medicine,¹⁰⁻¹² and agriculture.¹³⁻¹⁵ It is related to the cohesive forces between molecules, and it is responsible for some consequent phenomena such as capillarity, the shape of droplets and the behavior of interfaces with liquids. Despite its widespread applications and importance, predicting the SFT of organic compounds is a difficult task. This is due to the intricate chemical structure and varied properties of organic compounds.¹⁶

SFT largely varies among organic compounds, influencing their behavior during oil extraction, bubble formation, adhesion of liquids to solid surfaces, and the formulation of chemical products.¹⁷ Traditional techniques

for measuring SFT have been in use since the 1930s, with methods like the capillary rise method.¹⁸ Further research established a correlation between theoretical and experimental values of SFT,¹⁹ and approaches based on the generalized van der Waals theory presented high accuracy for experimental values of SFT measured for simple polar fluids.²⁰ Later, a theoretical model was developed²¹ based on density functional theory (DFT) and Barker-Henderson perturbation theory.²² With this model, the authors demonstrated that the proposed state equation could predict SFT with a deviation of 3.3%. More recently, classical molecular dynamics simulations were used to obtain interfacial tensions of oil,²³ which were then used to train a machine learning model which could predict the interfacial tensions with a minimum error of 2%.

Machine learning (ML) techniques have been applied to predict physical and chemical properties with great accuracy.²⁴ These techniques use sophisticated mathematical models that can learn from experimental data and further classify or predict outcomes for new instances. The capacity of these models was shown in recent articles that accurately

*e-mail: amaurijp@gmail.com; james.almeida@ilum.cnpem.br

Editor handled this article: André Galembeck (Guest)

Esteemed professor Oswaldo Luiz Alves, thank you for your invaluable guidance and support.



predicted an array of hydrocarbon properties such as boiling points, densities, and vapor pressures.²⁵ Another approach used a hybrid model that combined ML models such as linear regression and neural networks, with the aim of achieving a predictive model to obtain the SFT of hydrocarbons surfactants in aqueous media.²⁶ In addition, convolutional neural networks (CNN) were used to estimate the SFT using images of hanging drops of liquids, and the trained models had accuracy above 97%.²⁷ Recent efforts were taken to understand the features' significance in accurate ML models used to predict the SFT in binary and ternary mixtures containing ionic liquids (IL). In one case, authors used data sets containing IL and organic solvent parameters from universal *quasi*-chemical functional group activity coefficient model and Abraham solvation parameter model.²⁸ Another approach used a data set with features such as mole fraction, temperature, pressure, and types of molecules functional groups.²⁹

The development of accurate predictive models for SFT would allow the optimization of compound design, which is a key factor for improving the performance of any process involving fluid systems, especially hydrocarbon based. In this context, there is the premise that large and well cured data sets containing molecule characteristics (or features) such as thermodynamic state function values, intensive properties and phase transition information (e.g., melting and boiling points) can lead to ML predictors that successfully predict SFT. In the current state of these models, it is important to assess how features contribute to the success of the prediction accuracy. This optimization would determine which are the necessary experimental characterization of molecules to be taken for generating suitable data sets to feed ML models, thus largely reducing laboratory work and costs. In this study, we extend the investigation on the significance of features in predictive ML models for SFT. For this, we used extensive databases containing hydrocarbon physicochemical properties. These properties include enthalpy of vaporization at boiling point, density, thermal expansion coefficient, enthalpy of fusion, dipole moment, van der Waals (vdW) area and volume, radius of gyration, and isothermal compressibility. Our chosen models for analysis encompass linear regression (LR), K-nearest neighbor (KNN), random forest (RF), and XGBoost (extreme gradient boosting). By using the abovementioned properties, our purpose was to optimize the models to achieve accurate regressors for SFT. Finally, we evaluated the feature importance in the models with the methods of permutation feature importance (PFI) method and Shapley analysis (SHAP). The determination of feature importance is crucial to understand which molecular properties are necessary for the model to accurately predict the SFT.

Methodology

Data acquisition

We have obtained physical properties of hydrocarbons from Yaws.³⁰ Details on the number of molecules (instances) available in the handbook are shown in Table 1. To extract the data from Yaws' text, Tabula³¹ was used to convert the PDF into CSV files. This allowed data manipulation using Python libraries such as Pandas^{32,33} and Numpy,³⁴ which facilitate working with tabular data. To analyze and extract information from the molecular formula, two libraries, mol-mass (Molmass PyPI) and chemparse (Chemparse PyPI),³⁵ were utilized. This allowed the incorporation of information on atomic composition and molecular mass into the data sets.

Table 1. Number of instances present in the handbook³⁰ used to obtain the physicochemical properties of the molecules

Chapter/feature	No. of molecules
Enthalpy of vaporization at boiling	21337
Density solid	13781
Critical properties	9927
Density liquid	9766
Surface tension	9766
Thermal expansion coefficient	9766
Enthalpy of vaporization	8646
Enthalpy of fusion	5846
Dipole moment	1454
vdW area and volume	1411
Radius gyration	1103
Isothermal compressibility	650

vdW: van der Waals.

Data preparation

To ensure data consistency, certain conditions were used during the data cross-referencing process, such as having a common molecular identifier (the CAS number) in all tables and discarding non-standard information. Additionally, only SFT calculated at a temperature of 298.15 K were included. With these premises, we assembled four data sets combining different physicochemical properties obtained from the extraction of the handbook (see Table 2). Upon inspection of the last row and the distribution of the table, it becomes apparent that the four data sets are not uniform. They differ in the terms of the measured properties (features) and the number of molecules included (instances). As can be concluded by comparing data sets 0 and 3, the more completeness in terms of features (number of physicochemical properties considered) present in the data

set the less number of instances (molecules). Therefore, it is natural to investigate which data set significantly contributes to the prediction of the SFT in hydrocarbons.

Machine learning models

The methodology depicted in Figure 1 describes the approach taken to address the task of SFT prediction (\hat{y}) using a set of physicochemical properties (features) $x_{\text{feats}} \in \mathbb{R}^n$. This procedure is applied to all available data sets consistently. Considering the planned methodology, it is crucial to assess the dimensions of each data set to prevent the “curse of dimensionality”.³⁶ The dimensions of the data sets are as follows: data set 0 consists of 497 rows and 22 columns, data set 1 has 820 rows and 20 columns, data set 2 contains 3377 rows and 20 columns, and data set 3 comprises 4170 rows and 21 columns. The goal of the proposed work is to identify the data set that yields accurate predictions of \hat{y} using the ML-based regression algorithm. Although addressing correlations can be challenging in ML algorithms, it can be mitigated by transforming the features space to a lower-dimensional space while preserving high variance. In our approach, we employed a principal component analysis based (PCA-based) mapping technique to determine its contribution to the regression process. It is important to note that there is a trade-off when applying

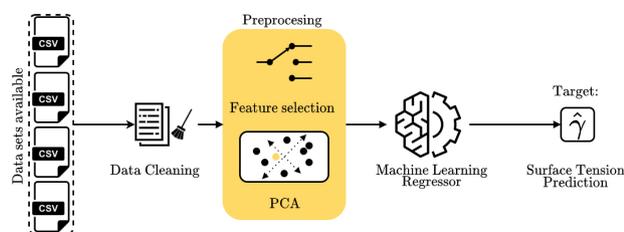


Figure 1. Diagram summarizing the approach used to obtain machine learning models capable of predicting SFT from tabulated physicochemical properties of molecules.

such techniques, as the interpretability of the original data set can be impaired in favor of the mapping process.

To determine the optimal regressor that yields the best performance with the most representative data set, it is necessary to establish a performance measurement criterion for the algorithms. Since the problem at hand involves regression, two metrics were utilized to assess algorithm performance: mean squared error (MSE) and the coefficient of determination (R^2) score. Therefore, the best regressor for the available data sets will exhibit the lowest MSE and the highest R^2 score. To accomplish this, a straightforward routine was applied, wherein each data set is loaded individually and the same procedure is applied to all of them to ensure a fair comparison among the techniques. In this case, four algorithms are tested to address the regression: (i) LR, (ii) KNN, (iii) RF and (iv) XGBoost.

The choice of these algorithms is based on their potential for generalization. LR is used as a baseline for comparing the performance of other algorithms. A LR model is a statistical technique used to establish linear relationships between dependent variables. More complex algorithms may demonstrate improved metrics if LR has sufficient generalization capability for the prediction task. The selection of other algorithms is based on their performance with tabular data, with RF and XGBoost being particularly effective according to the state of the art.³⁷ In a decision tree algorithm, data records are structured hierarchically, comprising nodes and branches guided by specific rules, rendering them suitable for classification and numerical (regression) data sets.³⁸ The RF algorithm extends the decision tree algorithm by creating and growing multiple decision trees for evaluation purposes.³⁹ Conversely, XGBoost is a scalable and distributed model based on gradient-boosted decision trees, featuring parallel tree boosting capabilities.³⁷ KNN was included as an intermediate choice to provide additional performance comparisons. This

Table 2. Data sets generated after data extraction (from Yaws handbook)³⁰ and after data crossing

Measurement	Data set 0	Data set 1	Data set 2	Data set 3
Surface tension	9766	9766	9766	9766
Density liquid	9766	9766	9766	9766
Enthalpy of vaporization at boiling	21337	21337	21337	21337
Enthalpy of vaporization	8646	8646	8646	8646
Enthalpy of fusion	5846	5846	5846	–
Thermal expansion coefficient	9766	9766	9766	9766
Dipole moment	1454	–	–	–
Radius of gyration	1103	–	–	–
vdW area and volume	1411	1411	–	–
Number of molecules	498	820	3377	4170

vdW: van der Waals.

algorithm is commonly used for classification and regression tasks, relying on the majority class or average value of the 'k' nearest neighbors in the feature space to make predictions for new data points. It operates without assumptions about data distribution, directly learning from the training data by measuring distances and selecting the nearest neighbors based on a specified value of "k".⁴⁰

Regarding the data distribution, each available data set is divided into two subsets: one exclusively for training the ML algorithms (85%) and another for testing the models generated by the algorithms (15%). A grid search is conducted for each model to explore a set of hyperparameters. Additionally, k-fold cross-validation is performed on a portion of the training data, ensuring accurate selection of the best model. In simplified terms, a set of potential regressors is obtained and those with the best performance on data not used for training is selected, considering metrics, hyperparameters, and the data set representativeness, these results are shown in the Supplementary Information (SI) section. Once the best regressor is chosen, the model is tested. This process is then repeated for each algorithm.

Models are often considered black boxes, making it challenging to directly understand how each property influences the predictions. In this context, we used the permutation feature importance method (PFI) and the SHAP (Shapley additive explanations)^{41,42} analysis to create detailed explanations of the model and to provide a clearer interpretation on how the features influence the target values. PFI assesses the importance of each feature in a machine learning model by observing how much the performance of the model drops when the values of a feature are randomly rearranged. Features that lead to significant drops in performance upon shuffling are considered more influential for the predictive ability of the model. On the other hand, the SHAP algorithm calculates Shapley values,⁴³ indicating the impact of each feature on the prediction. Negative SHAP values indicate an influence towards a lower target value and positive values indicate an influence towards a higher target. Additionally, Shapley values are useful in generating feature importance indicators. The relevance of applying this method lies in its ability to generate knowledge that can be applied even without the model, playing a crucial role, especially in the design of new materials.

Results and Discussion

PCA analysis on the features influence in the data sets

The primary objective of PCA is to keep a significant amount of information from the original data while

reducing its dimensionality. However, this reduction comes at the cost of losing variability in the data set. Figure S1 (SI section) demonstrates that by adding components to the data set the variance is largely increased, while the first component of each data set captures the highest variance. However, to preserve a substantial amount of variance (ca. 0.99), nearly all components are required. This poses a challenge for applying PCA since the reduction in dimensionality comes at the expense of losing interpretability of the original data sets. Since the dimensionality remains nearly the same, training an ML algorithm based on this representation would be redundant. Therefore, this representation is discarded as it loses interpretability while achieving a dimensionality similar to that of the original data, as illustrated in Figure S1 (SI section).

ML performance evaluation

The outcomes obtained from the data preprocessing and training of the ML algorithms are presented in Table 3. It is evident that the performance of each algorithm varies across the different data sets, with data set 2 yielding the best results for all implemented techniques. Among the algorithms, XGBoost (XGB) demonstrates the highest performance, indicating superior generalization capabilities on unseen data during training. This is reflected in the R^2 and MSE metrics (see Table 3). It is interesting to note that data set 0 has more features, including the dipole moment, radius of gyration, and vdW area and volume. These features could be important in the prediction of the SFT. On the other hand, the low performance of models for this data set is possibly related to its small size. The same can be observed for the data set 1 (it retained the vdW area and volume), which is the second-smallest data set. For the data sets 2 and 3, the only difference among them is the enthalpy of fusion (present in data set 2). Although the removal of enthalpy of fusion in data set 3 lowers the number of molecules in the data set, it was observed that this feature has a positive effect on the performance of the model. Hence, there is a balance between the size of the data set and the information provided by its features. One can also see from Table 3 that the XGBoost regressor performed better than all the other methods in almost all cases.

To observe the performance of each technique on unseen data, a random subset of the data is partitioned multiple times. The experiment is repeated 100 times to generate Figure 2, which displays the boxplots for each algorithm and the corresponding metrics. The focus is solely on data set 2, as it consistently produced the best

Table 3. Overall ML algorithms performance for all data sets considering MSE and R² values

Technique		Data set 0	Data set 1	Data set 2	Data set 3
LR	MSE / (mN ² m ⁻²)	642.68	168.26	11.59	19.70
	R ²	0.31	0.20	0.71	0.74
KNN	MSE / (mN ² m ⁻²)	769.37	14.27	6.62	15.74
	R ²	0.17	0.81	0.83	0.79
RF	MSE / (mN ² m ⁻²)	671.48	10.29	5.46	15.99
	R ²	0.28	0.86	0.87	0.80
XGBoost	MSE / (mN ² m ⁻²)	707.91	16.05	4.65	11.32
	R ²	0.24	0.78	0.89	0.85

LR: linear regression; KNN: K-nearest neighbor; RF: random forest; XGBoost: extreme gradient boosting; MSE: mean squared error; R²: xxx.

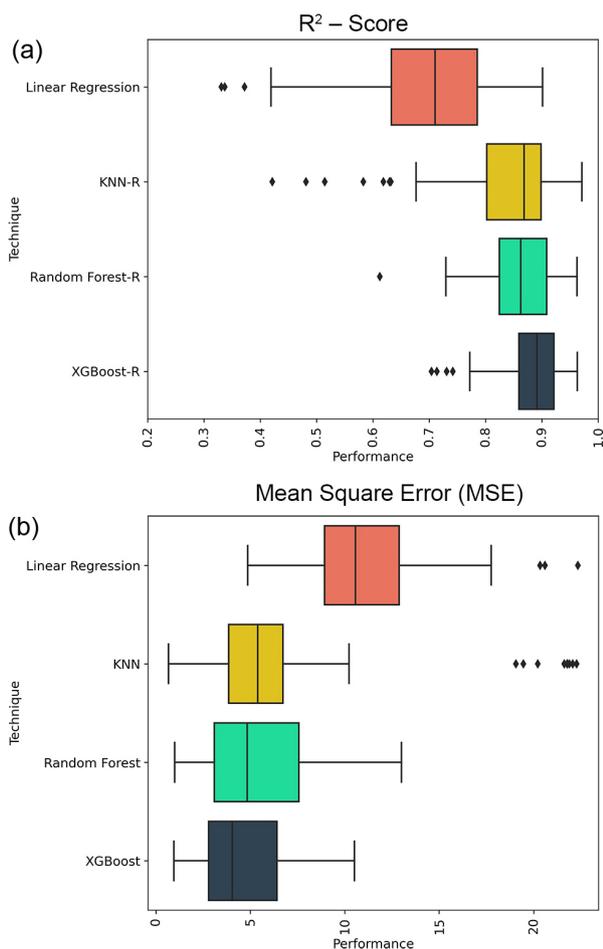


Figure 2. Box plot depicting the performance of each model trained with data set 2. Performance was analyzed in terms of (a) the R² score and (b) mean squared error (MSE). The diamonds depicted in the boxplots represent outliers.

results with all ML algorithms, as previously demonstrated. Notably, XGBoost (XGB) exhibits superior performance compared to the other algorithms in both metrics. It maintains a lower error level (approximately 4.5 mN² m⁻²; see Figure 2b), while achieving the highest prediction score (approximately 0.89; see Figure 2a). The results presented in Table 3 align with the boxplots shown in Figure 2, as the

values in the table deviate slightly from the median in the boxplots, but remain close to the expected values.

Although the metrics already indicate the ability of the trained model to generalize, further interpretability is desired, especially considering that dimensionality reduction was not performed. To gain insights into which features are relevant for predicting SFT (\hat{y}), we analyzed the feature importance with the PFI method. Since XGBoost (XGB) was the best performing classifier, we focus on reporting the importance analysis for this algorithm trained on data set 2. Figure 3 illustrates that the most important features for XGBoost are the enthalpy of vaporization, density, molecular weight, and the number of oxygen atoms in the molecule (O). In a more detailed interpretation of this result, we understand that features like density can be easily correlated with the SFT, since this latter can increase with an increasing density value due to intermolecular forces and interactions (although there is not a universal trend for this relation). However, the expected importance of the molecule hydrogen (H) and carbon (C) content in the prediction process was not

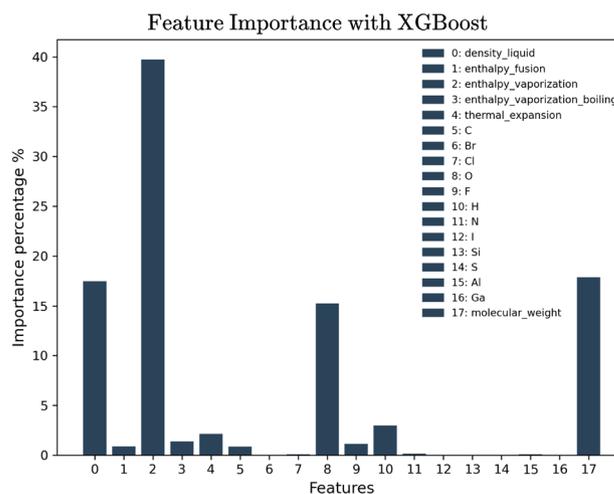


Figure 3. Feature importance obtained with the PFI method for the XGBoost algorithm, measured as the percentage of importance in the prediction process of the target variable \hat{y} (using data set 2).

substantial. One possible explanation at the model level is that these two variables are highly correlated, which is expected given that hydrocarbons predominantly consist of these two atoms, and usually most of the hydrogens are bonded to carbon atoms.

It is important to note that not all trained algorithms exhibit the same features with the same level of relevance. We analyzed the feature importance for the other algorithms as well as for data set 3. Figure S2 (SI section) showcases the most relevant features for algorithms that typically rely on decision trees, particularly RF and XGBoost. It can be observed the significance of the density as a relevant parameter in both regressors for all data sets. However, Figure S2 highlights that the importance of density varies across the models and data sets. Additionally, there are other features that are consistently shared among all models, such as the number of hydrogen atoms in the molecule (H), enthalpy of vaporization and molecular weight.

The feature importance obtained with the Shapley (SHAP) analysis for XGBoost algorithm on data set 2 is shown in Figure 4. The enthalpy of vaporization, density, molecular weight and hydrogen content are among the most important features (see Figure 4). The molecule oxygen content (O) is more critical for the PFI than for the SHAP analysis. The order of the feature importance also changed among the two methods. Considering the SHAP values for enthalpy of vaporization the lower the value the lower the interfacial tension (see Figure 5). This could be explained by the fact that a lower enthalpy of vaporization implies that the molecules have weaker intermolecular interactions at the surface, thus requiring less energy to be separated.⁴⁴ The same trend is observed for the lower values of density, as the molecule tends to have lower interfacial tensions (although with some exceptions).⁴⁵ This happens because denser liquids tend to require some energy to alter its surface area, i.e., molecules can be more closely packed, which, in turn, means stronger intermolecular forces.⁴⁶ For the molecular weight, it is commonly known that long chain surfactants can have lower interfacial tensions.⁴⁷ The carbon and hydrogen content (C and H) also shows lower SFT for higher contents (see Figure 5). This can be related to the size of the hydrocarbon chain, as longer chains have more carbon/hydrogen atoms and present smaller SFT values.

Just with four features (enthalpy of vaporization, density, molecular weight and the molecule hydrogen content) the model achieved an explainability of approximately 80% (see Figure 4). All the other features should play a secondary role in determining the SFT. Features like thermal expansion, oxygen amount (O), enthalpy of fusion and all other 9 features (see Figure 3)

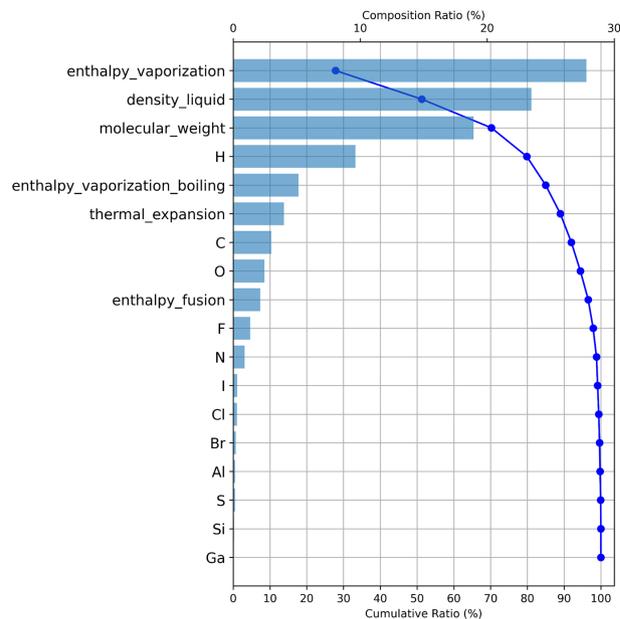


Figure 4. Feature importance using the Shapley (SHAP) analysis for the XGBoost algorithm trained with data set 2. The upper x-axis represents the composition ratio, and the lower x-axis represents the cumulative ratio. The bars are related to the composition ratio, and the line is related to the cumulative ratio.

did not manifest a clear correlation in SHAP results. For instance, an increase in the amount of oxygen in the molecule (O) is associated with both large and small values of SFT (see Figure 5), thus indicating a more complex correlation for this feature. The presence of oxygen is directly related to the manifestation of hydrogen bonds in the system, which, in turn, modifies the SFT. However, hydrogen bonds are also strongly related to the molecule stereochemistry. This is why this feature did not present an easily explainable SHAP result.

A limitation observed for the most accurate model here described (XGBoost trained on data set 2) was the prediction of SFT in regions of the features space containing little data density. In these regions, we found a larger error for the model, as seen in Figure 6. For the test set described in Figure 6 we used 507 molecules (instances). Possibly, by increasing the number of molecules in the set we should observe a better accuracy of the model in these regions (of low data density).

Recent efforts were made to create predictive models for SFT. Using the simplified molecular input line entry specification (SMILES), researchers were able to predict SFT with great accuracy (R^2 of 0.98-0.99) using ML-based models trained with 244 molecules.⁴⁸ Another article reported ML models capable of predicting 15 physicochemical properties for 23 fuel types, including alkanes, alkenes, alcohols, aldehydes, ketones, esters, ethers, aromatics, peroxide and carboxylic acids.⁴⁹ The authors

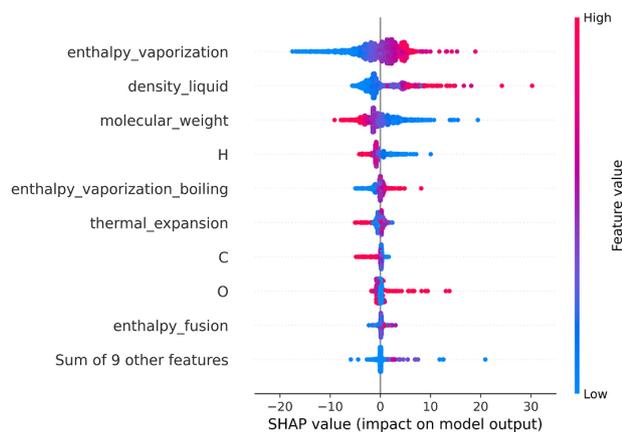


Figure 5. Beeswarm plot of SHAP values. The color bar represents the feature value, and the x-axis is the SHAP value. The points in the graph comprise all data used to train the model (using data set 2).

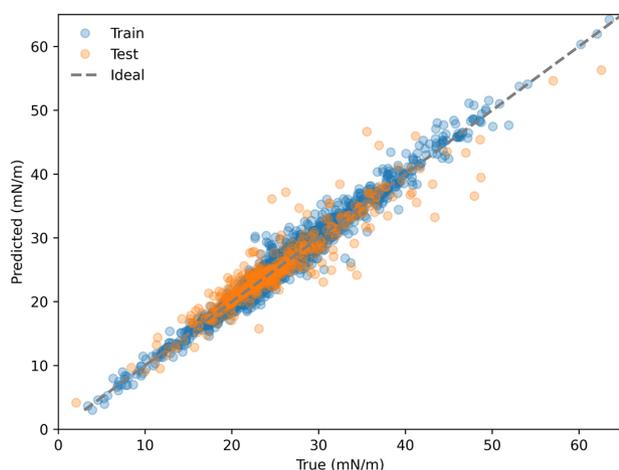


Figure 6. Graph showing the prediction accuracy of the XGBoost model (trained with data set 2) for training (Train; blue dots) and test (Test; red dots) sets. The x-axis represents the true values, and the y-axis represents the predicted values (each dot represents a data point). The closer the dots are to the diagonal dashed line the more accurate the predictions are.

used quantitative structure property relationship (QSPR) as input to the models and achieved accuracies of 0.9898 (R^2) and 0.7993 mN m^{-1} (RMSE, root-mean-square error). However, they used just 488 molecules as instances. In another recent article,²⁶ a group of 154 model hydrocarbon surfactants had their SFT predicted with an accuracy of 0.69-0.87 (R^2 value). Although high accuracies were previously obtained in SFT predictive ML models, the generalization capacity of these models remains to be evaluated. To the best of our knowledge, the models presented here were trained with larger data sets of hydrocarbon molecules. In this sense, their generalization capacity (considering unseen molecules) was more tested in comparison with the previous ones reported in the literature. Just for comparison, in the testing set of our models we have used more molecules than the whole training data set of previous articles (> 500 molecules).

Conclusions

Following a thorough exploration of various machine learning algorithms for predicting hydrocarbon surface tension (SFT), we observed that XGBoost has superior performance in terms of predictive accuracy and generalization capabilities compared to other ML algorithms. This conclusion was taken after exploring data sets with different sizes and number of features, and a balance was found among these two properties of the data sets. The most accurate machine learning model was XGBoost (trained with 3377 molecules), with a mean squared error (MSE) of 4.65 mN m^{-1} and an R^2 score of 0.89. The models also indicate that the SFT prediction can be accurately done using enthalpy of vaporization, density, molecular weight and hydrogen content as data input. In this sense, this work provides a guide for selecting the most important molecule characterization experiments to be performed to feed predictive machine learning models for SFT.

In conclusion, this study presents a promising framework for predicting hydrocarbon SFT, which holds potential applications in diverse fields, including the oil and gas industry and materials science. Further exploration and refinement of these predictive models can contribute to advancements in various practical domains. While our discoveries shed light on the molecular properties that impact SFT, this work leaves the perspective for additional investigation into other molecular properties.

Supplementary Information

Supplementary information (models hyperparameters, PCA results and feature importance analysis for all data sets) is available free of charge at <http://jbc.ssbq.org.br> as PDF file.

Acknowledgments

The authors acknowledge the National Council for Scientific and Technological Development (CNPq) for the INCT Materials Informatics grant and CCM-UFABC for the computational resources. AJP acknowledges Ceará State Research Funding Agency (FUNCAP) for grant PRONEM PNE-0112-000480100/16.

Author Contributions

Paula J. F. Cala was responsible for data curation, investigation, visualization, writing original draft; Guilherme G. Dariani for data curation, investigation, visualization, writing original draft; Eduardo

T. A. Veiga for data curation, investigation, visualization, writing original draft; Pedro Henrique D. de Macedo for data curation, investigation, visualization, writing original draft; Amauri J. Paula for conceptualization, formal analysis funding acquisition, investigation, project administration, writing (original draft, review and editing); James M. Almeida for conceptualization, formal analysis funding acquisition, investigation, project administration, writing (original draft, review and editing).

References

- Sheng, J. J.; *Adv. Pet. Explor. Dev.* **2013**, *6*, 1. [Link] accessed in June 2024
- Massarweh, O.; Abushaikha, A. S.; *Energy Rep.* **2020**, *6*, 3150. [Crossref]
- Tavakkoli, O.; Kamyab, H.; Shariati, M.; Mohamed, A. M.; Junin, R.; *Fuel* **2022**, *312*, 122867. [Crossref]
- Baret, J.-C.; *Lab Chip* **2012**, *12*, 422. [Crossref]
- Lee, W.; Walker, L. M.; Anna, S. L.; *Macromol. Mater. Eng.* **2011**, *296*, 203. [Crossref]
- Cox, M. F.; *J. Am. Oil Chem. Soc.* **1986**, *63*, 559. [Crossref]
- Moharram, K.; Abd-Elhady, M.; Kandil, H.; El-Sherif, H.; *Energy Convers. Manage.* **2013**, *68*, 266. [Crossref]
- Chelazzi, D.; Bordes, R.; Giorgi, R.; Holmberg, K.; Baglioni, P.; *Curr. Opin. Colloid Interface Sci.* **2020**, *45*, 108. [Crossref]
- Pironti, C.; Motta, O.; Ricciardi, M.; Camin, F.; Cucciniello, R.; Proto, A.; *Talanta* **2020**, *219*, 121256. [Crossref]
- Maldonado-Valderrama, J.; Wilde, P.; Macierzanka, A.; Mackie, A.; *Adv. Colloid Interface Sci.* **2011**, *165*, 36. [Crossref]
- Percival, S. L.; Mayer, D.; Malone, M.; Swanson, T.; Gibson, D.; Schultz, G.; *J. Wound Care* **2017**, *26*, 680. [Crossref]
- Mc Callion, O.; Taylor, K.; Thomas, M.; Taylor, A.; *Int. J. Pharm.* **1996**, *129*, 123. [Crossref]
- Paria, S.; *Adv. Colloid Interface Sci.* **2008**, *138*, 24. [Crossref]
- del Carmen Hernández-Soriano, M.; Degryse, F.; Smolders, E.; *Environ. Pollut.* **2011**, *159*, 809. [Crossref]
- Hernández-Soriano, M. C.; Peña, A.; Mingorance, M. D.; *J. Environ. Qual.* **2010**, *39*, 1298. [Crossref]
- Wu, X.; Huang, J.; Zhuang, Y.; Liu, Y.; Yang, J.; Ouyang, H.; Han, X.; *Appl. Sci.* **2023**, *13*, 4200. [Crossref]
- Katz, D. L.; Saltman, W.; *Ind. Eng. Chem. Res.* **1939**, *31*, 91. [Crossref]
- Swartz, C. A.; *Physics* **1931**, *1*, 245. [Crossref]
- Almeida, B. S.; Telo da Gama, M. M.; *J. Phys. Chem.* **1989**, *93*, 4132. [Crossref]
- Abbas, S.; Nordholm, S.; *J. Colloid Interface Sci.* **1995**, *174*, 264. [Crossref]
- Fu, D.; Lu, J.-F.; Liu, J.-C.; Li, Y.-G.; *Chem. Eng. Sci.* **2001**, *56*, 6989. [Crossref]
- Barker, J. A.; Henderson, D.; *J. Phys. Chem.* **1967**, *47*, 2856. [Crossref]
- Kirch, A.; Celaschi, Y. M.; de Almeida, J. M.; Miranda, C. R.; *ACS Appl. Mater. Interfaces* **2020**, *12*, 15837. [Crossref]
- Zhou, Q.; Lu, S.; Wu, Y.; Wang, J.; *J. Phys. Chem. Lett.* **2020**, *11*, 3920. [Crossref]
- Dobbelaere, M. R.; Ureel, Y.; Vermeire, F. H.; Tomme, L.; Stevens, C. V.; Van Geem, K. M.; *Ind. Eng. Chem. Res.* **2022**, *61*, 8581. [Crossref]
- Seddon, D.; Müller, E. A.; Cabral, J. T.; *J. Colloid Interface Sci.* **2022**, *625*, 328. [Crossref]
- Soori, T.; Rassoulinejad-Mousavi, S. M.; Zhang, L.; Rokoni, A.; Sun, Y.; *Fluid Phase Equilib.* **2021**, *538*, 113012. [Crossref]
- Deng, J.; Zhang, Y.; Jia, G.; *Phys. Fluids* **2023**, *35*, 062101. [Crossref]
- Lei, Y.; Shu, Y.; Liu, X.; Liu, X.; Wu, X.; Chen, Y.; *J. Taiwan Inst. Chem. Eng.* **2023**, *151*, 105140. [Crossref]
- Yaws, C. L.; *Thermophysical Properties of Chemicals and Hydrocarbons*, 1st ed.; Elsevier: London, UK, 2008.
- Aristarán, M.; Tigas, M.; Merrill, J. B.; <https://tabula.technology/>, accessed in June 2024.
- Pandas*, 2.2.2; NumFOCUS Foundation, USA, 2009.
- Pandas*, <https://pandas.pydata.org>, accessed on June 23, 2024.
- Harris, C. R.; Millman, K. J.; van der Walt, S. J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N. J.; Kern, R.; Picus, M.; Hoyer, S.; van Kerkwijk, M. H.; Brett, M.; Haldane, A.; del Río, J. F.; Wiebe, M.; Peterson, P.; Gérard-Marchant, P.; Sheppard, K.; Reddy, T.; Weckesser, W.; Abbasi, H.; Gohlke, C.; Oliphant, T. E.; *Nature* **2020**, *585*, 357. [Crossref]
- Boyer, G.; Ignatenko, V.; *Chemparse*, 0.3.1, 2024.
- Bellman, R. E.; *Dynamic Programming*; Dover Publications: Mineola, USA, 2003.
- Chen, T.; Guestrin, C.; *XGBoost*, 2.1.0, 2024.
- Kamali, M. Z.; Davoodi, S.; Ghorbani, H.; Wood, D. A.; Mohamadian, N.; Lajmorak, S.; Rukavishnikov, V. S.; Taherizade, F.; Band, S. S.; *Mar. Pet. Geol.* **2022**, *139*, 105597. [Crossref]
- Barjouei, H. S.; Ghorbani, H.; Mohamadian, N.; Wood, D. A.; Davoodi, S.; Moghadasi, J.; Saberi, H.; *J. Pet. Explor. Prod. Technol.* **2021**, *11*, 1233. [Crossref]
- Fix, E.; Hodges Jr., J. L.; *Int. Stat. Rev.* **1989**, *57*, 238. [Crossref]
- Lundberg, S. M.; Lee, S.-I. In *Advances in Neural Information Processing Systems*, vol. 30; Curran Associates, Inc.: Long Beach, CA, USA, 2017.
- Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I.; *Nat. Mach. Intell.* **2020**, *2*, 56. [Crossref]
- Shapley, L. S. In *Contributions to the Theory of Games (AM-28)*, vol. 2; Kuhn, H. W.; Tucker, A. W., eds.; Princeton University Press: Princeton, NJ, USA, 1953, p. 307-318.
- Alibakhshi, A.; *Fluid Phase Equilib.* **2017**, *432*, 62. [Crossref]

45. Jia, L.; Ma, X.; Shi, B.; *J. Macromol. Sci., Part B: Phys.* **2010**, *50*, 376. [Crossref]
46. Bitaab, A.; Taghikhani, V.; Ghotbi, C.; Ayatollahi, S.; *J. Chem. Thermodyn.* **2008**, *40*, 1131. [Crossref]
47. Smit, B.; Schlijper, A. G.; Rupert, L. A. M.; Van Os, N. M.; *J. Phys. Chem.* **1990**, *94*, 6933. [Crossref]
48. Goussard, V.; Duprat, F.; Ploix, J.-L.; Dreyfus, G.; Nardello-Rataj, V.; Aubry, J.-M.; *J. Chem. Inf. Model.* **2020**, *60*, 2012. [Crossref]
49. Li, R.; Herreros, J. M.; Tsolakis, A.; Yang, W.; *Fuel* **2021**, *304*, 121437. [Crossref]

Submitted: February 7, 2024

Published online: June 26, 2024