A logo for college computing

Description automatically generated

**Assessment Cover Page**

|  |  |
| --- | --- |
| *Student Full Name* | Daniela Mariano Barreto |
| *Student Number* | 2023278 |
| *Module Title* | Strategic Thinking |
| *Assessment Title* | CA 3 Final Submission |
| *Assessment Due Date* | 11th May 2024 |
| *Date of Submission* | 11th May 2024 |

**Declaration**

By submitting this assessment, I confirm that I have read the CCT policy on academic misconduct and understand the implications of submitting work that is not my own or does not appropriately reference material taken from a third party or other source.

I declare it to be my own work and that all material from third parties has been appropriately referenced.

I further confirm that this work has not previously been submitted for assessment by myself or someone else in CCT College Dublin or any other higher education institution.

**Applied machine learning to estimate material to adsorb CO2**

by

*Daniela Mariano Barreto*

*Higher Diploma in Science in Data Analytics for Business Strategic Thinking*

*James Garza*

*CCT College*

*Dublin, Ireland*

Abstract

200 – 250 WORDS

[NOTE: This section is designated for the abstract. Abstracts are not assigned page numbers and should precede the table of contents. If an abstract is unnecessary for your work, please delete this page.]

Attention: All notes must be removed from the document before submission!!

**Key-words:** k-nearest neighbours (k-NN), Gradient Boosting (GB), CRISP-DM, ….

GitHub link:

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

Contents

[**Introduction** 1](#_Toc165282434)

[**Business description** 2](#_Toc165282435)

[*Hypothesis* 2](#_Toc165282436)

[*General goal* 2](#_Toc165282437)

[*Scope* 3](#_Toc165282438)

[*Success criteria/indicators* 3](#_Toc165282439)

[**Technologies** 4](#_Toc165282440)

[*Libraries* 4](#_Toc165282441)

[*Models and machine learning algorithms* 4](#_Toc165282442)

[*Hyperparameters tuning and cross validation* 4](#_Toc165282443)

[**Accomplishment** 4](#_Toc165282444)

[*Data* 4](#_Toc165282445)

[*Machine learning models* 4](#_Toc165282446)

[**Challenges** 4](#_Toc165282447)

[*Problem definition* 4](#_Toc165282448)

[**Cross Industry Standard Process - Data Mining (CRISP-DM)** 5](#_Toc165282449)

[Business understanding 6](#_Toc165282450)

[Data understanding 6](#_Toc165282451)

[Data preparation 11](#_Toc165282452)

[Machine learning models screening 12](#_Toc165282453)

[Machine learning implementation 12](#_Toc165282454)

[**Evaluation** 13](#_Toc165282455)

[**Conclusions** 16](#_Toc165282456)

[**Future recommendations** 16](#_Toc165282457)

[**Project's timeline overview** 16](#_Toc165282458)

[**Ethical considerations** 17](#_Toc165282459)

[**Appendices** 17](#_Toc165282460)

[**References** 20](#_Toc165282461)

[**Supplementary information** 22](#_Toc165282462)

[**References for supplementary information** 28](#_Toc165282463)

# **Introduction**

Concern about the increase in carbon dioxide (CO2) in the atmosphere began decades ago. Still, this problem continues nowadays with unsatisfactory predictions for 2040, in which an increase of approximately 70% and 25% is estimated to produce electricity and coal-based, respectively (Huetteman, Bowman and Slater-Thompson, 2016, p.81; Ren and Liu, 2023, p.1).

Studies show that CO2 emissions have caused enormous global damage due to the greenhouse effect, a potential explanation for climate change. These changes probably cause a decrease in biodiversity and harm to human lives due to tsunamis and earthquakes, for example; consequently, a drastic decrease in ecosystem productivity and economic disparities (Daneshvar *et al*., 2022, p.1; Ren and Liu, 2023, p.1).

Considering all, we see why mitigating the removal of CO2 has been the focus of several institutions, and the increase in the development of technologies and products is the central reflection of these concerns and actions that have increasingly gained momentum. (Ren and Liu, 2023, p.1). Asif *et al*. (2018, p.4) and Daneshvar *et al*. (2022, p.6) have shown the general panorama of all techniques that we have so far (Figure 1.)

A diagram of a structure

Description automatically generated

**Figure 1:** Techniques for CO2 capture.

This capstone project proposes to focus on the adsorption technique, considering the multitude of options available within each type of technology. This promising method has gained notable attention due to its low operational cost, lower energy demand, ease of handling, and general reliability (Daneshvar *et al*., 2022, p.4).

In light of the advancements in technologies for CO2 capture, we propose the implementation of machine learning (ML) models to determine the most suitable materials for adsorbing a desired amount of CO2 for a company. Each material possesses its own peculiarities, necessitating specific action plans that can subsequently be refined to lower production costs and enhance economic feasibility.

# **Business description**

In today's world, many industries are seriously considering developing sustainable business practices in response to the escalating consequences of increased CO2 emissions. Additionally, consumers increasingly favour companies with a 'green' stamp of approval.

In light of this, it is considered a fictitious company specializing in services and materials for CO2 capture and intends to modernize its systems by implementing machine learning models. By analysing data from academic articles, the ML model aims to identify materials most effective for CO2 adsorption, tailored to the specific needs of each sector. This approach could potentially personalize solutions, optimizing the efficiency of CO2 capture in various industries. It could also make significant contributions to environmental sustainability.

## *Hypothesis*

The question raised for this project is whether it will be possible to predict the most suitable material, considering a given CO2 concentration that needs to be captured.

**H0:** There is no significant difference in the CO2 adsorption capacity among the materials tested.

**H1:** There is a significant difference in the CO2 adsorption capacity among the materials tested.

## *General goal*

Principal objective:

The main objective of this project is to assess the performance of various machine learning models to find the best one capable of identifying the most suitable material for capturing CO2, considering a predetermined amount of the gas. Since each industry emits different concentrations of CO2, they may be interested in implementing certain materials to adsorb this pollutant. Thus, the model aims to estimate which material would best suit the concentration of CO2 needed for adsorption by the industry.

Secondary objectives:

* Identify the principal materials used in the adsorption technique.
* Use the Exploratory Data Analysis (EDA) to understand the data.
* Apply machine learning models to choose one with a high metric score.
* Implement the chosen ML model to predict a material given a specific amount of CO2.

Below is a scheme illustrating the objectives and workflow of this project.

**Figure 2:**

## *Scope*

This project aimed to assess several machine learning models to estimate which materials are suitable for adsorbing a certain concentration of CO2. Several articles were analysed, and the records that met the minimum criteria were compiled into the dataset, which was firstly analysed using descriptive and inferential statistics.

Subsequently, three machine learning models— Decision Tree Classifier (CART), Gradient Boosting Classifier (GBM), and Random Forest Classifier (RF)—were implemented for the core analysis. To address challenges like class imbalance, SMOTE was used. Random search and stratified K-Fold Cross-Validation were employed to adjust hyperparameters.

The project's scope was carefully defined, emphasizing the exclusive use of data derived from adsorption techniques within the past fifteen years. Moreover, a rigorous selection criterion was applied to source academic papers from reputable journals, ensuring data credibility. Temperature and pressure conditions were standardized to 25°C and 1 bar, respectively, to minimize external interference and ensure impartial results.

## *Success criteria/indicators*

# **Technologies**

## *Libraries*

## *Models and machine learning algorithms*

## *Hyperparameters tuning and cross validation*

# **Accomplishment**

## *Data*

## *Machine learning models*

# **Challenges**

Including strategies used to overcome them.

## *Problem definition*

Considering that the emission of CO2 is a global problem, government and private institutions are investing in research to find solutions in this field. Consequently, we expect to have a lot of data, however, we might face difficulties finding articles with comparable experimental conditions. The second problem might be in choosing the most suitable ML model.

# **Cross Industry Standard Process - Data Mining (CRISP-DM)**

CRISP-DM has been used as a project management methodology (figure 2a) with the breakdown of the phases. The project plan (figure 2b) is from gathering data to a conclusion, and all stages are interchangeable; the evaluation can generate results that can return to the preprocessing phase, for example.

The action plan will search for data with comparable experimental conditions, such as pressure and temperature. Use Python in Jupyter Notebook to clean and organize the data and create visualizations for an overview (EDA). Perform descriptive statistics and preprocessing the data if necessary.

Posteriorly, experiment with some ML models to obtain better results using cross-validation and hyperparameters adjustment. Evaluate the model’s performance and return to any previous phases if necessary. At the end, conclude the findings.



**Figure 2:** a) CRISP-DP scheme. Source: <https://healthdataminer.com/wp-content/uploads/2019/11/800px-CRISP-DM_Process_Diagram.png>. b) Project plan.

## Business understanding

The focus on sustainability has led industries to request technologies to capture CO2 from their emissions. This study aims to estimate which material should be suitable in relation to a given concentration of CO2 that the industry wishes to adsorb. This is important to provide information to create a specific plan that meets the sector's needs, avoiding unnecessary expenses.

## Data understanding

*Data source*

Data were gathered from an article that compiled hundreds of sources, focusing on three extensively studied materials: carbon-based materials, metal-organic frameworks (MOFs), and polymers (Dziejarski et al., 2023, pp. 3, 20, 36, 41). However, during the previous phase of this project, the need for additional data arose. While I found more records related to carbon-based materials (Yuan et al., 2021, p.11927), there were no comparable findings for the other materials. This lack of comparable findings is consistent with Dziejarski et al. (2023, p.3), which indicates that most of the research is focused on carbon-based materials. All the data used are presented in the Supplementary Information section in Table S1, followed by references to the sources of each record.

*Data description*

The dataset is composed of 11 features, and 92 records with 3 classes of material used in the CO2 adsorption (carbon-based, metal-organic frameworks (MOFs), and microporous polymers). Features description in table 1.

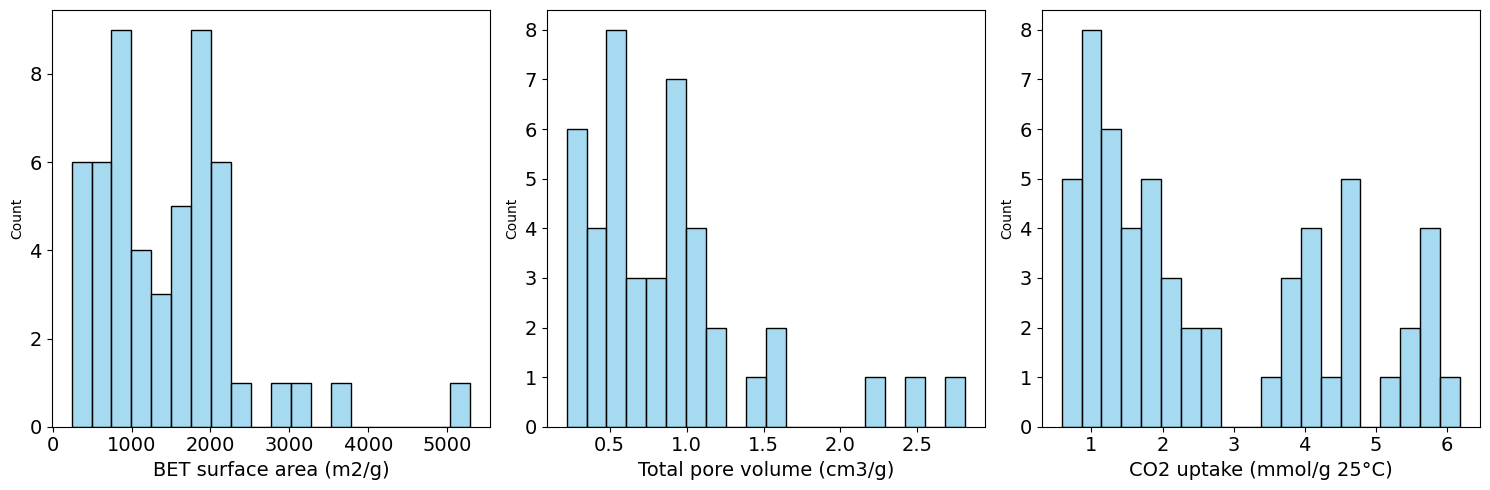
**Table 1:** Data dictionary. A white paper with black text

Description automatically generated

*Exploratory data analysis (EDA)*

For this analysis, I focused on these features: 'Material', 'BET surface area\_m2/g' (SA), 'Total pore volume\_cm3/g' (TPV), and 'CO2 uptake\_mmol/g\_25°C' (CO2\_uptake), because they are common to all types of materials and might be related to the CO2 adsorption.

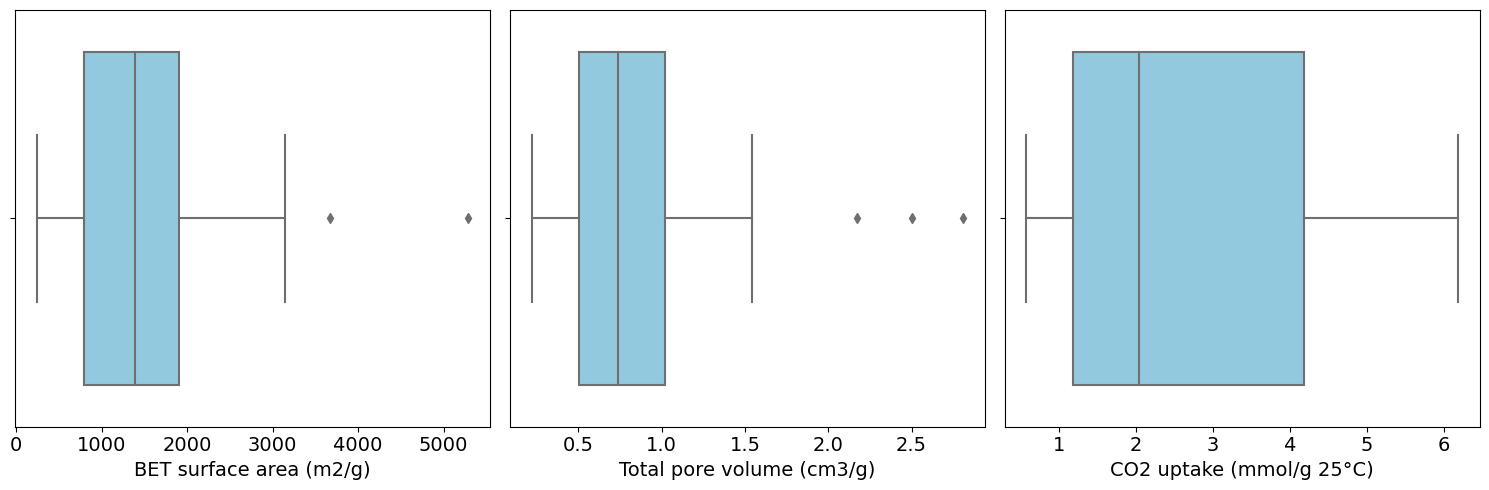
Due to the presence of missing values in two features (SA and TPV), it was necessary to impute them. But first, it was essential to evaluate the descriptive statistics, which revealed that the data was skewed, indicating that the data in these two columns do not follow a normal distribution. This information was crucial for selecting the appropriate method to impute the missing values, such as mean, median, or interpolation. In this case, in which the features were skewed and had outliers, imputation with the median tends to be a reasonable approach, as interpolation is indicated when the data change over time, which is not the case here. Moreover, the median is generally more robust to outliers. REF



a

e

c



d

b

f

**Figure 3:** Histogram and boxplot before data imputation. 'BET surface area\_m2/g' (a and b), 'Total pore volume\_cm3/g' (c and d), ‘CO2 uptake\_mmol/g\_25°C’ (e and f).

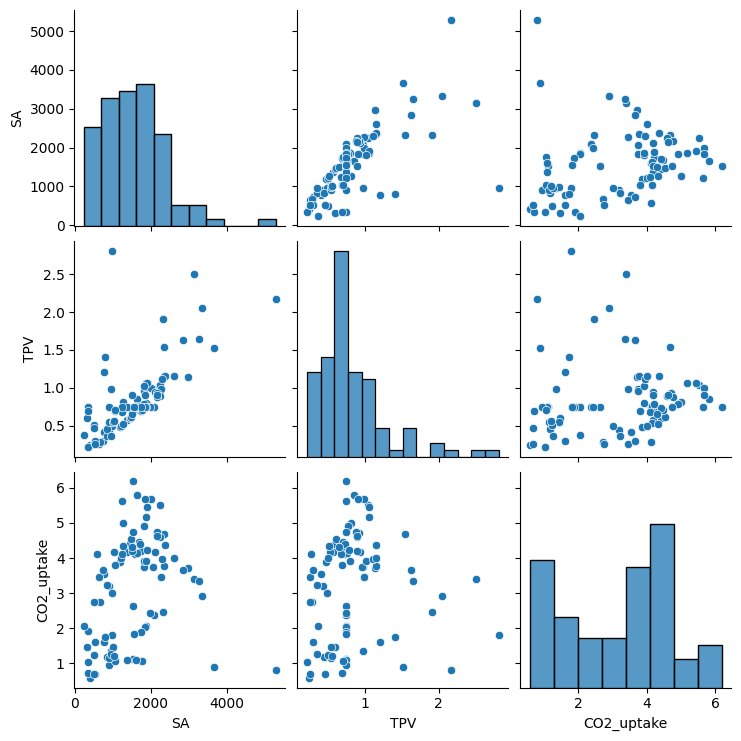
*Data Cleaning*

The dataset had no duplicates, but 6.5% of the values were outliers. Therefore, since it was a significant amount, I decided not to remove them because it could compromise the data integrity, especially considering the presence of imbalanced classes that might affect the model's performance. To address this issue, I will seek a robust machine learning method to handle it.

*Descriptive Statistics*

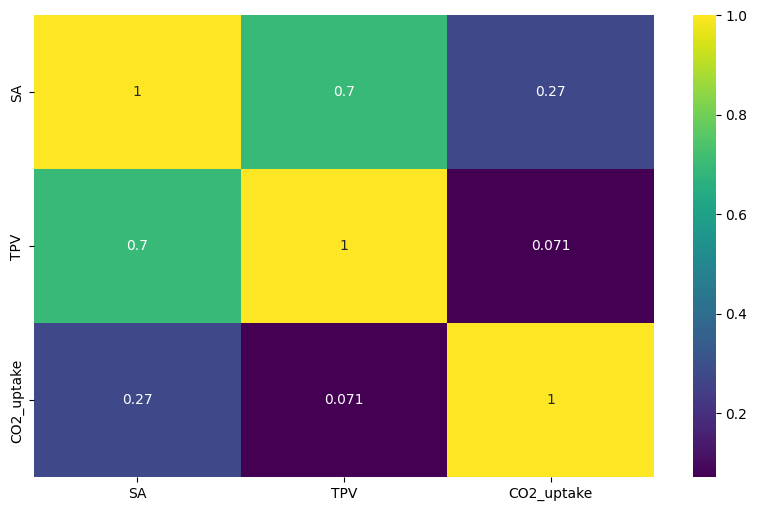
Summary statistics parameters showed that data is skewed and is reasonable sparse due to the encode with zero the carbon-based material. The minimum and maximum values presented a considerable variance, and each feature's scale is very different; thus, it will be necessary to scale the data.

Figures 4 presented the pair plot and heatmap. The results showed a moderately strong positive correlation between SA and TPV (0.70); thus, there was a tendency to increase area at the same time as increasing volume and vice versa. However, when comparing CO2\_uptake with SA (0.23), the correlation was weak if compared with Area and Volume. In addition, practically, there was no correlation between CO2\_uptake and TPV (0.085), meaning little or no linear correlation that can be negligible.



a

b



**Figure 4:** a) Pair plot, and b) heatmap of features correlation. SA: BET surface area (m2/g). TPV: Total pore volume (cm3/g). CO2\_uptake: CO2 uptake (mmol/g 25°C).

*Inferential Statistics*

Considering that the dataset does not follow a normal distribution, I opted to apply the Kruskal-Wallis test because it tends to be appropriate when handling a non-normal distribution and small sample size, as it does not make assumptions about the data's normality (Devore, 2012, p.645).

With a 95% confidence level, I found that there is a significant difference in the CO2 adsorption capacity among the materials tested. Therefore, the material that can adsorb more CO2 in this case is Carbon-based, followed by MOFs and then Polymers (see figure 5).

This result can provide important information that can be used to tailor the best material to an industry's needs. For instance, if a company needs to adsorb 4 mmol/g of CO2 at 25°C and 1 bar, the model should directly suggest a 'Carbon-based' material, which could avoid expending time and resources on other materials that may not meet their requirements. Employing a machine learning model is expected to yield a more refined selection process.

A graph of different colored squares

Description automatically generated

**Figure 5:** CO2 uptake in three different materials under a pressure condition of 1 bar.

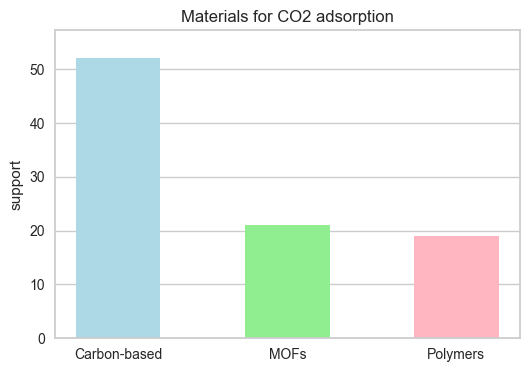
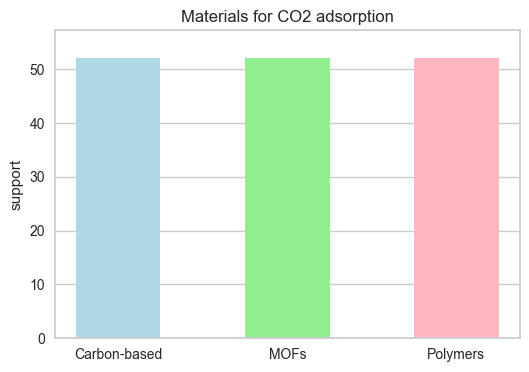
## Data preparation

*Preprocessing*

Data scaling is a common preprocessing step for many machine learning models because there is a considerable variation between features that can contribute to inaccurate predictions since features with larger scales may dominate the calculations (Müller and Guido, 2017, p.138). In this project, I employed the RobustScaler function, which is more suitable in the presence of outliers. This is essential for many ML models (scikit-learn, n.d.), especially considering there are outliers in this dataset.

Considering that the data have significantly imbalanced classes, I handled this challenging using the Synthetic Minority Over-sampling Technique (SMOTE), that potentially results in better generalization compared to traditional methods (Chawla, 2002; Chawla, 2010, p. 881; imbalanced-learn.org, n.d.) (Figure 6).

Additionally, I applied one-hot encoding to avoid attributing undue weight to categories, ensuring that all materials have equal importance.



a

b

**Figure 6:** a) Sample without SMOTE. b) Sample with SMOTE.

## Machine learning models screening

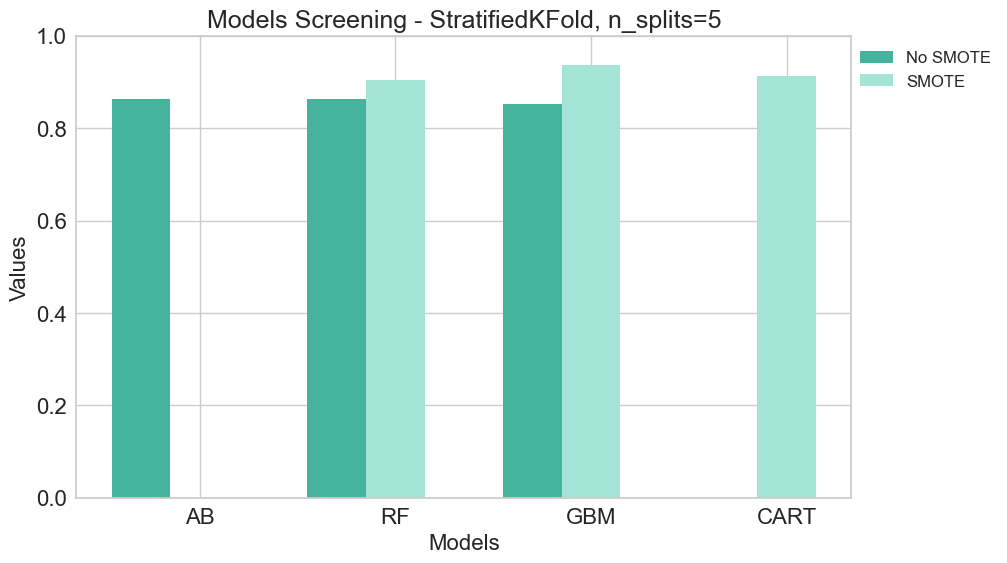
In this section, I experimented with seven ML models to find the best three to use in this project to predict the most suitable material to adsorb CO2, given the concentration of this gas.

The chosen metric is F1 -score when the data is imbalanced and accuracy when the data is balanced with SMOTE because accuracy might not be appropriate when the data is imbalanced; the F1-score metric tends to be more reliable in this case as it is a trade-off between precision and recall in the rare class (Chawla, 2010, p. 876-878). This differentiation is essential to reduce bias when giving more weight to one class over the other.

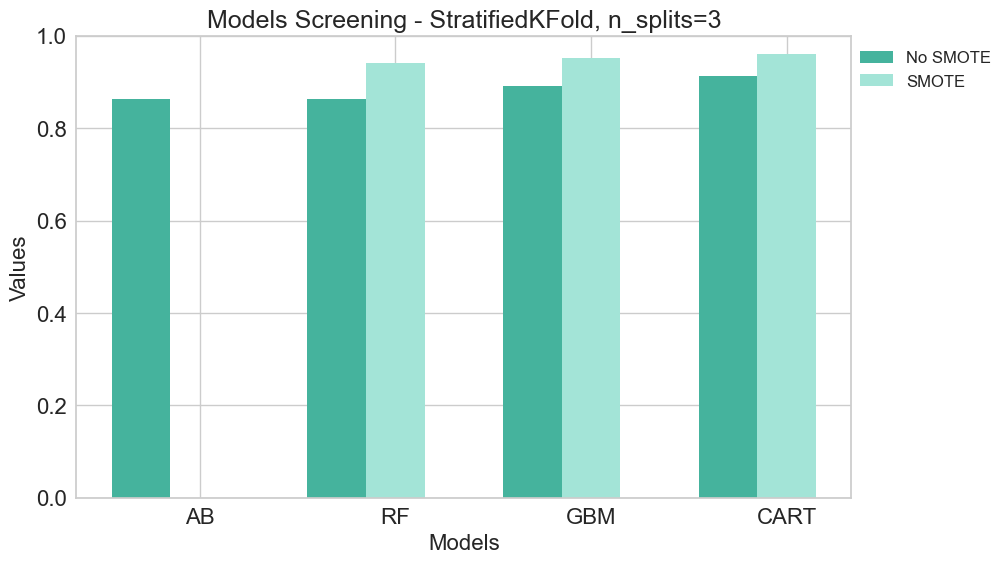
I also tested n\_splits equal to 5 and 3 using Stratified K-Fold, as this approach is usually recommended when dealing with imbalanced classes. It partitions the dataset into k folds while ensuring that each fold maintains the same class distribution as the original dataset (Prusty, Patnaik, and Dash, 2022, p.3), preserving the percentage of samples for each class, which is essential when dealing with imbalanced datasets.

Considering that in this project I am interested in prioritizing true positive results, I focused on identify models that correctly identify positive instances, because of this I evaluated metrics such as recall, precision, and F1-score. Looking at the charts (Figure 7), we see that all tested conditions performed well, with scores above 0.86. Also, the top models performed similarly across these conditions. However, the condition without SMOTE and n\_splits equal to 3 notably improved the precision, recall, and F1-score in the two minority material classes, MOFs (class 1) and Polymers (class 2). Given the significance of accurate classification for these minority classes in comparison to the majority class, Carbon-based (class 0), I will opt for these conditions to implement the three best-performing models, namely: Decision Tree Classifier (CART), Gradient Boosting Classifier (GBM), and Random Forest Classifier (RF) equally to Ada Boost Classifier (AB).

**Figure 7:** Machine learning model screening results: a) Stratified K-Fold with n\_splits=5 and b) Stratified K-Fold with n\_splits=3.

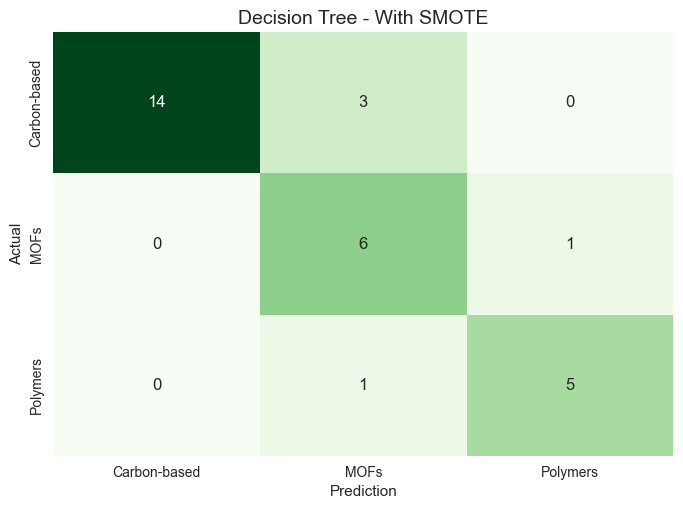
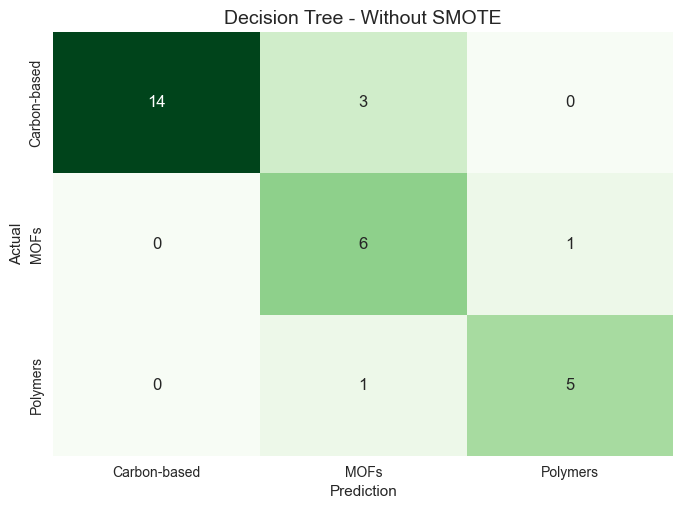


a



b

In the confusion matrix with SMOTE and without SMOTE using n\_splits = 3 (Figure 8), the misclassification for the minority classes (MOFs and Polymers) was only one instance for each class, a promising result.



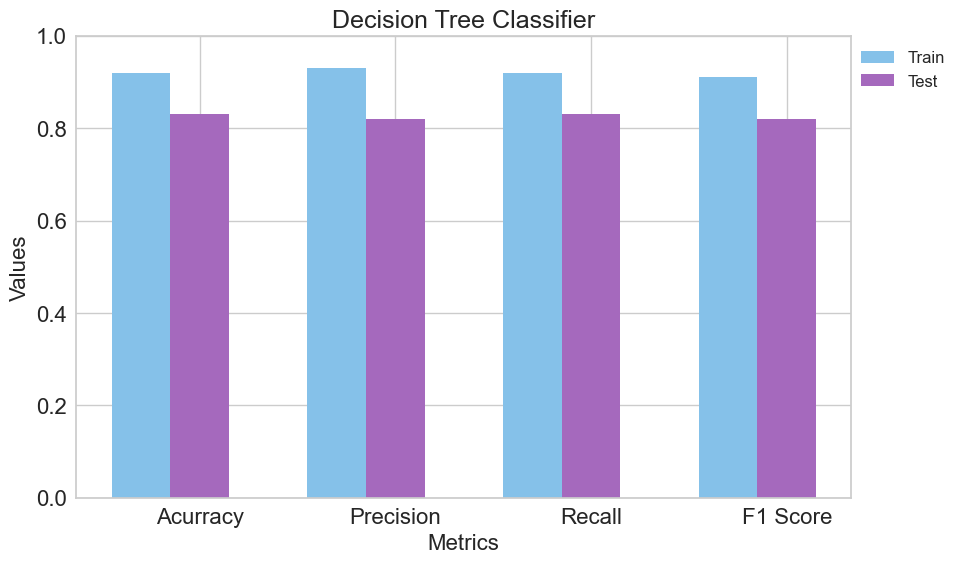
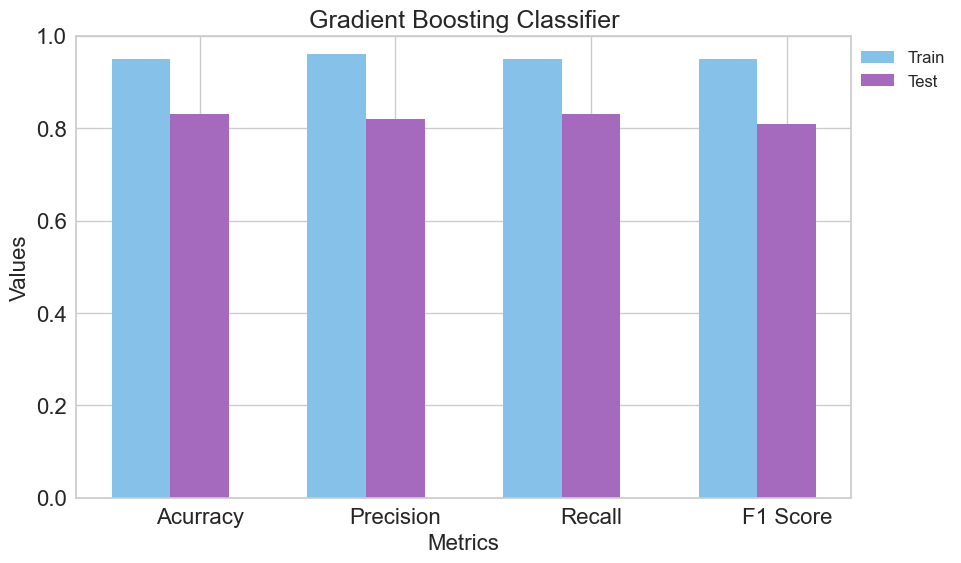
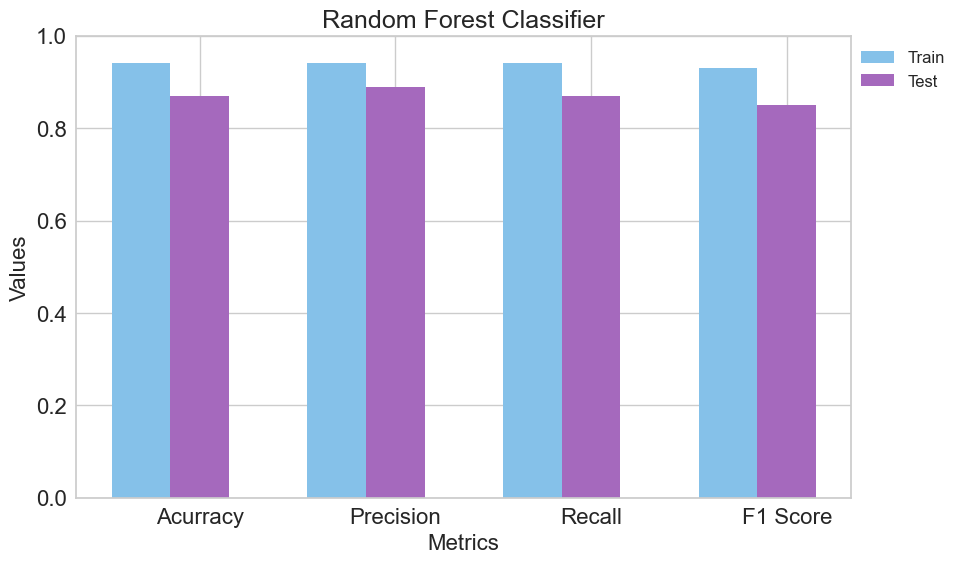
**Figure 8:** The confusion matrix of the model’s screening is on the left without SMOTE use and on the right with SMOTE use, both are in the test set.

## Machine learning implementation

The models selected in the previous section (CART, GBM, and RF) were implemented using Stratified K-Fold (n\_splits=3) to maintain class distribution in each fold, ensuring robust validation without data leakage, both with and without SMOTE. Employing Random Search Cross Validation facilitated hyperparameter optimization for each model, and their performance was evaluated using metrics such as accuracy, precision, recall, and F1-score. Additionally, I utilized a pipeline to set up the three models simultaneously, integrating both Stratified K-Fold and Random Search Cross Validation.

This project focuses on achieving high values for recall, as it is crucial to identify all positive samples, thus avoiding false negatives, while precision serves to limit the number of false positives, and the F1-score represents a trade-off between these two metrics (Müller and Guido, 2017, pp. 282-283).

In the experimentation without SMOTE, high values were found for all metrics in both the train and test sets (Figure 9). However, a significant difference exists between these sets in both Decision Tree and Gradient Boosting models, indicating that they may be slightly overfitting. The Random Forest model appears to generalize better to the test set in comparison with the other models. Additionally, the recall value was greater than the others, meaning that the Random Forest model is capable of correctly identifying true positive samples 87% of the time. These results prompted experimentation using SMOTE on the training set in an attempt to improve the models' ability to generalize.



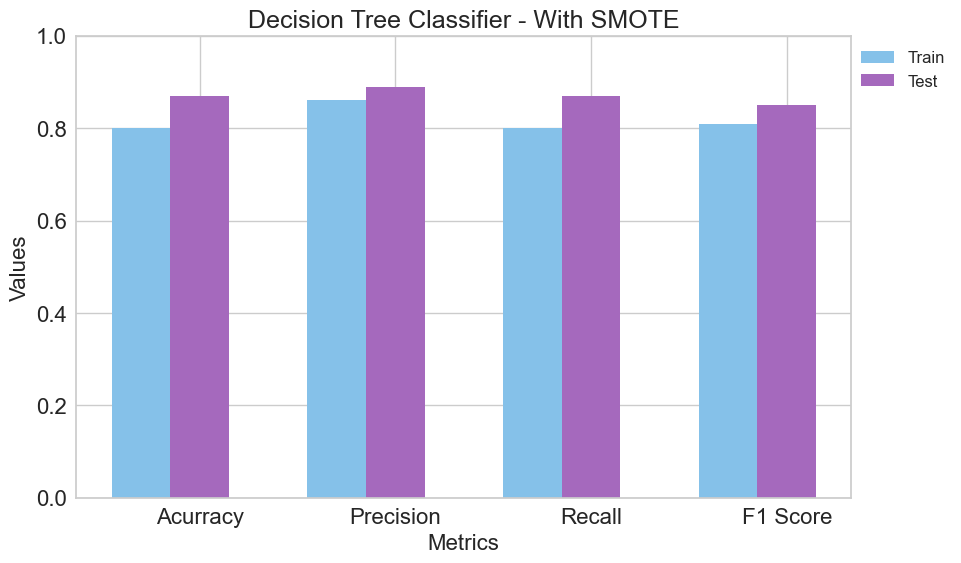
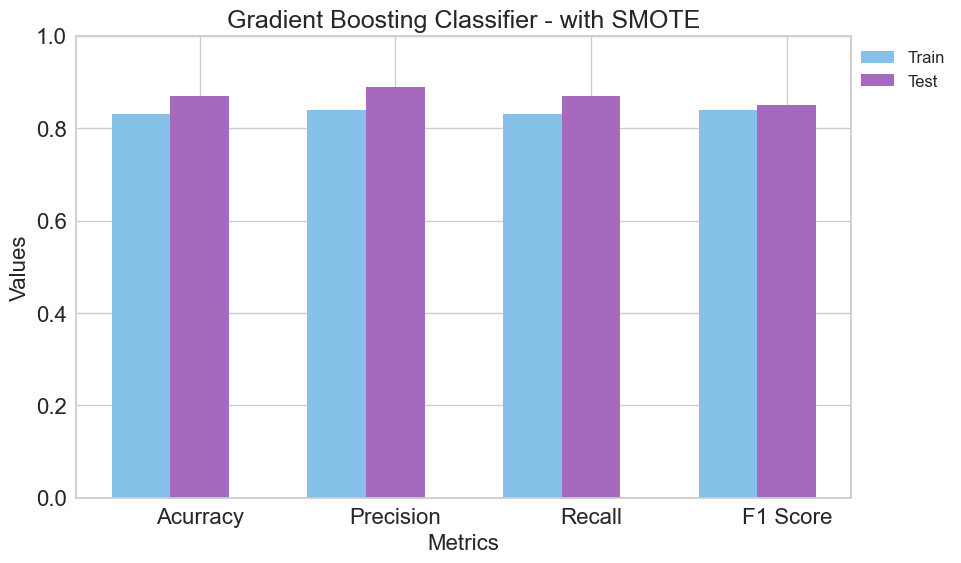
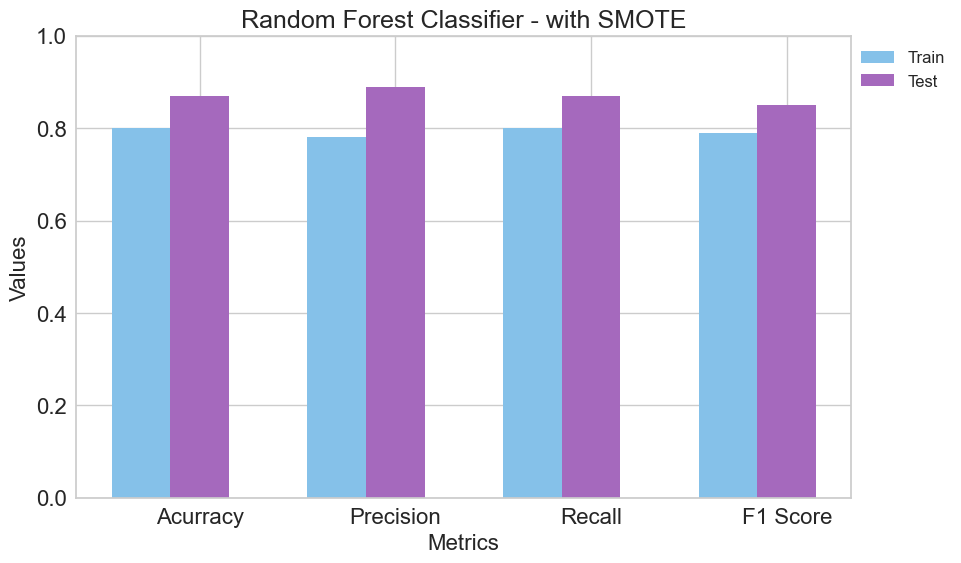
a

b

**Figure 9:** Models implementation without SMOTE. a) Decision Tree, b) Gradient Boosting, and c) Random Forest.

c

The use of SMOTE resulted in the improvement of the three models (Figure 10), which exhibited good performance in both the training and test sets, suggesting that there is not significant overfitting with any of these models. This indicates that the models can generalize well to unseen data. Furthermore, focusing on the recall metric—highlighting the accuracy in identifying actual positive samples—all models demonstrated similar results (ranging from 0.80 to 0.83) in the training set and 0.87 in the test set, reaffirming their strong ability to generalize. Moreover, Gradient Boosting exhibited greater consistency across other metrics, with negligible disparities observed between the datasets.



a

b

c

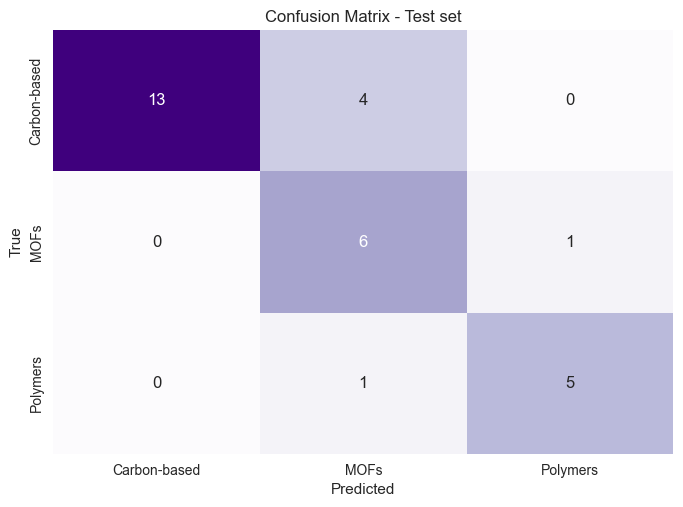
**Figure 10:** Models implementation with SMOTE. a) Decision Tree, b) Gradient Boosting, and c) Random Forest.

Considering the metrics of the three models, similar performance was observed in terms of true positive instances (recall) on the test set, with each model achieving a recall of 0.83. This indicates that 83% of the instances were correctly identified as positive by each model. Similarly, precision yielded consistent results, with 82% of the samples predicted as positive actually being positive.

The confusion matrix (Figure 11) of the training set resulted in correctly predicting true positives in all classes for all models, which generally indicates that the model is performing well when the dataset is balanced.

Upon further examination of the confusion matrix, it is evident that the test set excelled in identifying the majority class (Carbon-based). However, despite the enhancements brought about by SMOTE, both MOFs and Polymers still encounter challenges in minimizing false positives. This emphasizes the necessity of improving precision to address this issue effectively, although the precision score remains high (0.89) for the test set and all models. On a positive note, the Decision Tree model exhibited commendable performance in identifying the minority classes, with just one misclassification observed.

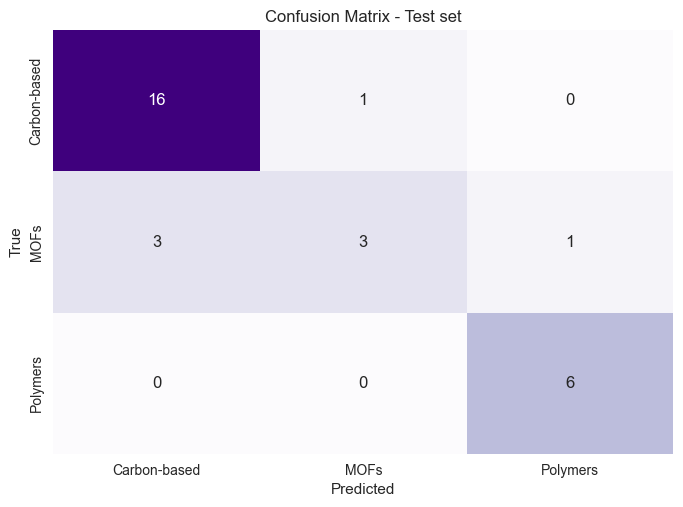
**Figure 11:** Confusion matrix with SMOTE. a) Decision Tree, b) Gradient Boosting, and c) Random Forest.



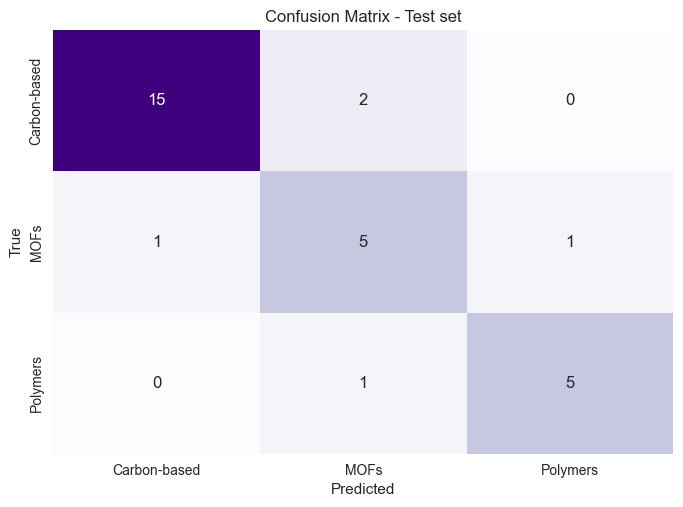
a



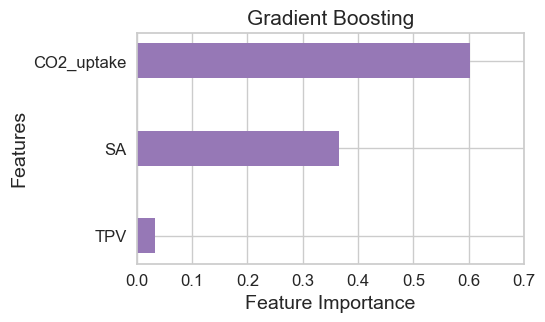
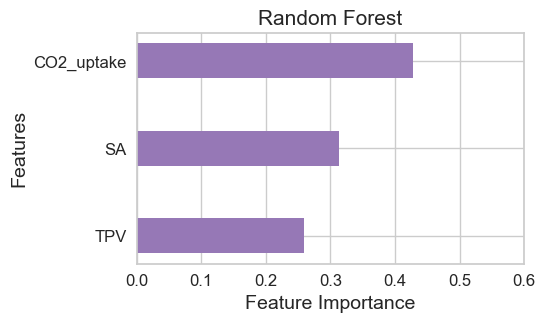
b



c

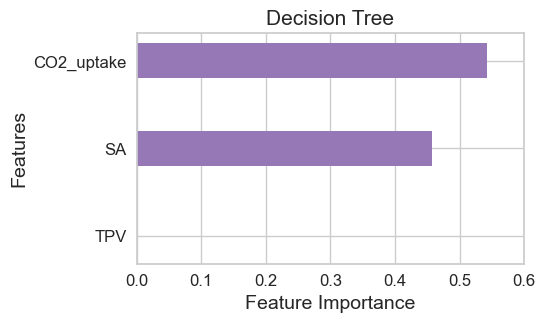


In Figure 12, the feature importance for the three models is presented. For all models, CO2\_uptake emerges as the most important feature, followed by SA and TPV. This suggests that the models' predictions were primarily influenced by CO2\_uptake and less influenced by TPV. Considering the moderately strong positive correlation between SA and TPV, as observed in the Descriptive Statistics section, removing the TPV feature could potentially improve predictions. However, since the model is already performing well, removing a feature may increase the risk of overfitting due to the simplification of the models. Therefore, I have decided not to remove it.



c

b



a

**Figure 12:** Feature Importance. a) Decision Tree, b) Gradient Boosting, and c) Random Forest.

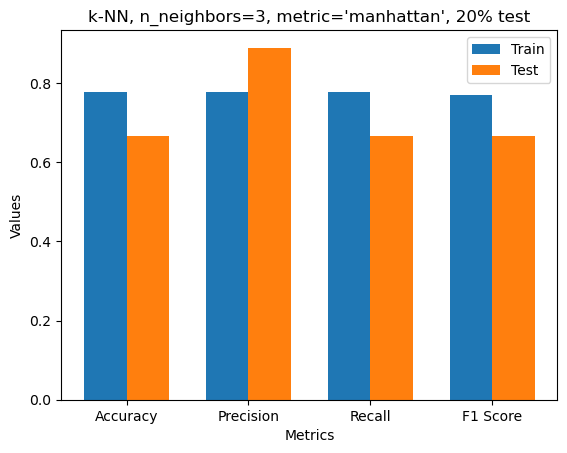
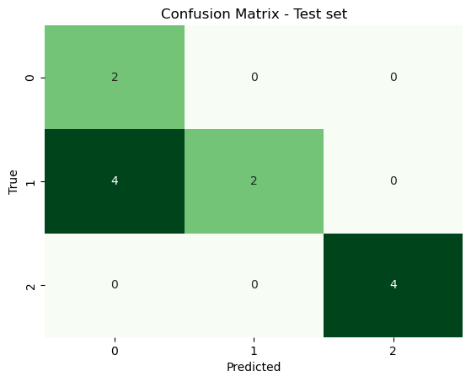
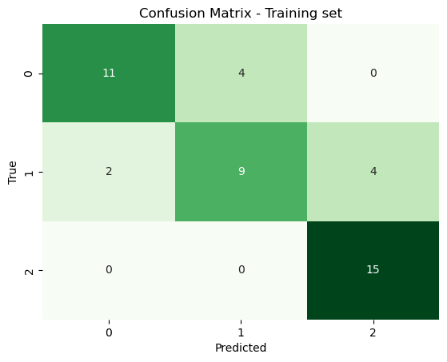
I applied the k-nearest neighbours (k-NN) because it is used with many classes and often performs reasonably without complex adjustments (Müller and Guido, 2017, p.36,44). In addition, I applied the Gradient Boosting (GB) model to improve the estimation, as it builds serial trees that try to correct the errors of the previous and provide better accuracy with the correct configuration (Müller and Guido, 2017, p .88,89). I also used cross-validation, manual hyperparameters adjustment, and Grid Search to achieve the optimal hyperparameters. SMOTE was used to deal with class imbalanced.

The target variable will be the 'Material\_replace' because I want to identify which material can be suitable given a certain concentration of CO2 uptake desired by the industry.

# **Evaluation**

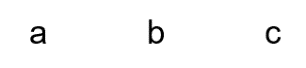
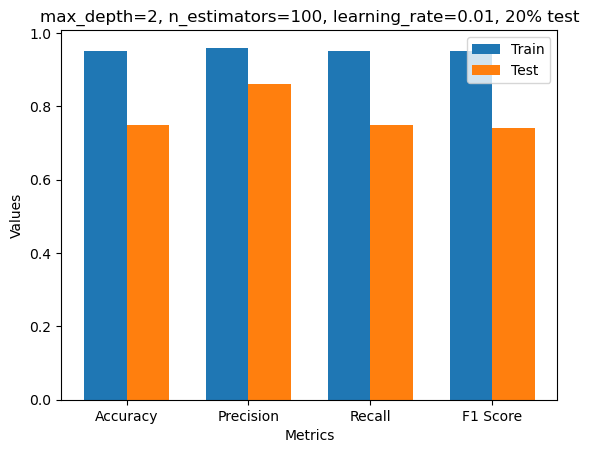
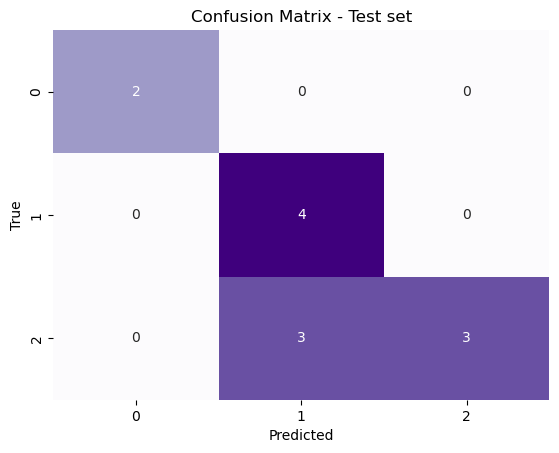
*Summary findings*

k-NN (figure 6) training and test sets presented reasonable performance, with better training results than the test, indicating that this model may be overfitting and incapable of generalizing well. SMOTE increased these differences (Jupyter Notebook file).



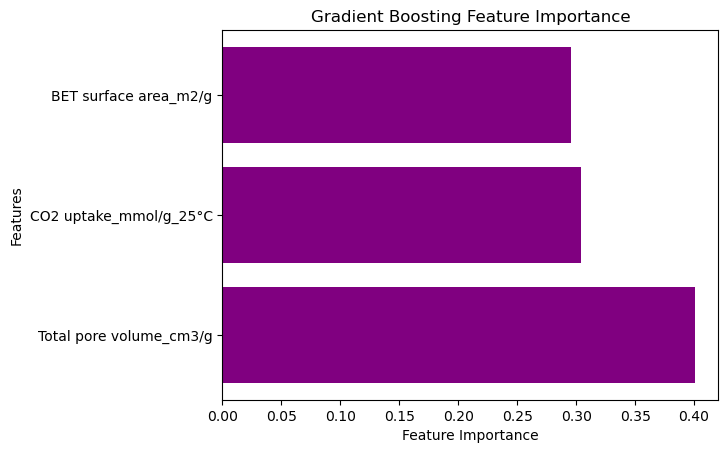
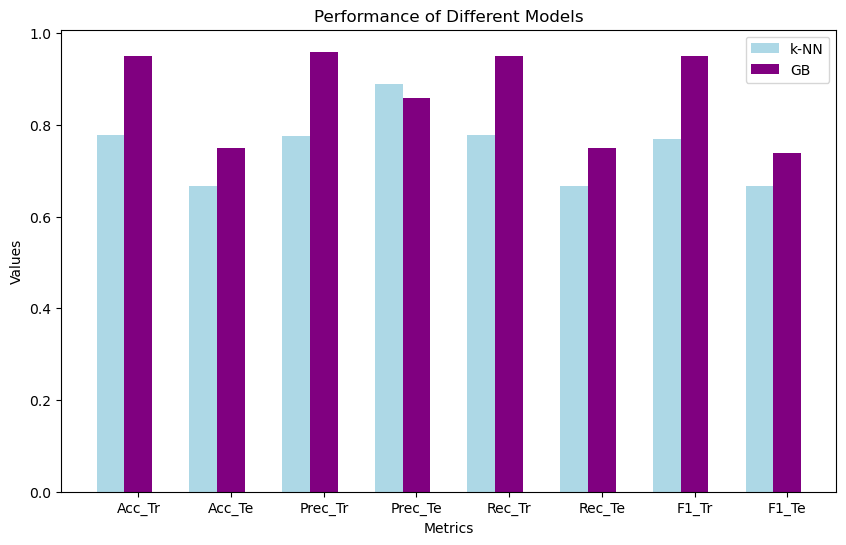
**Figure 6:** k-NN: Confusion matrix for training (a) and test (b) sets. C) Metrics for training and test sets.

GB (figure 7) improved compared to k-NN on both sets and although model training performed very well (probably overfitting), the test set may not generalize well to unseen data. Precision is high in both sets, indicating a high probability of correct prediction for positive instances in actual results, despite the test set capturing 75% of positive cases predicted (recall).



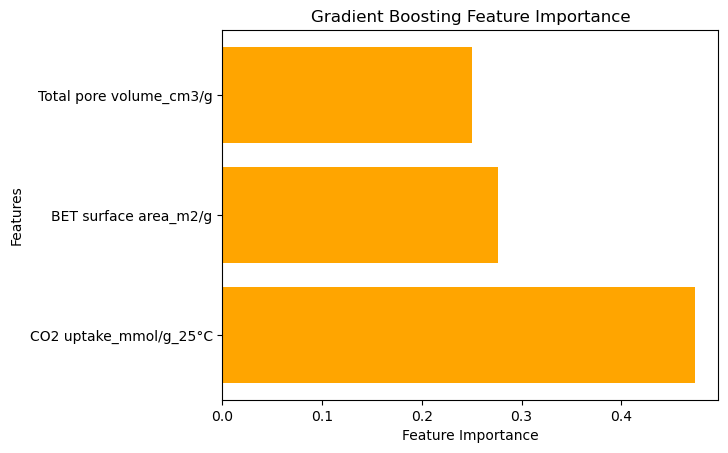
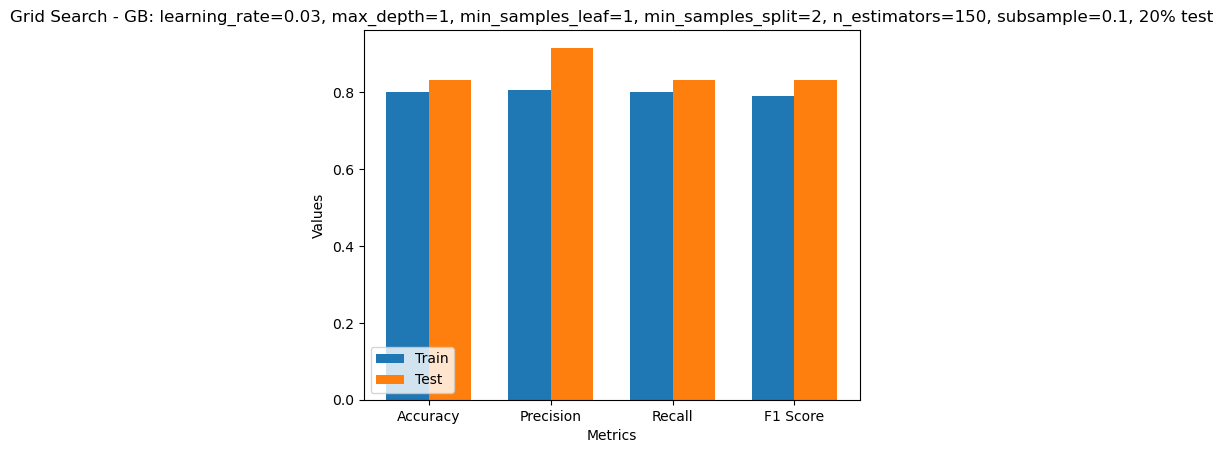
**Figure 7:** GB: Confusion matrix for training (a) and test (b) sets. C) Metrics for training and test sets.

Figure 8 summarises these two models to facilitate comparison. Neither generalised well to new data and may result in overfitting. In addition, the most important feature of the GB is 'Total pore volume', followed by the 'CO2 uptake'.



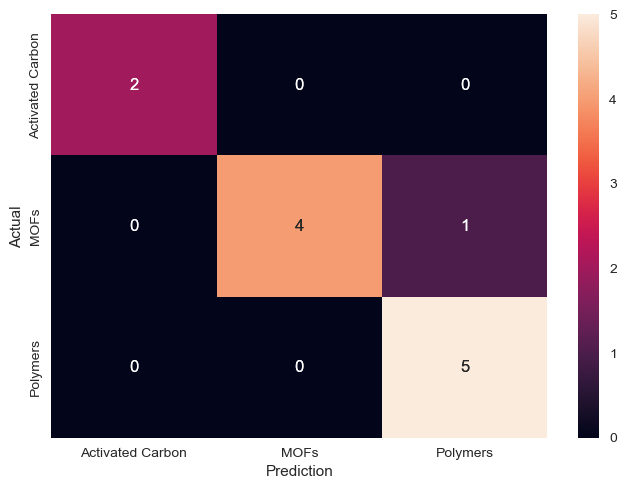
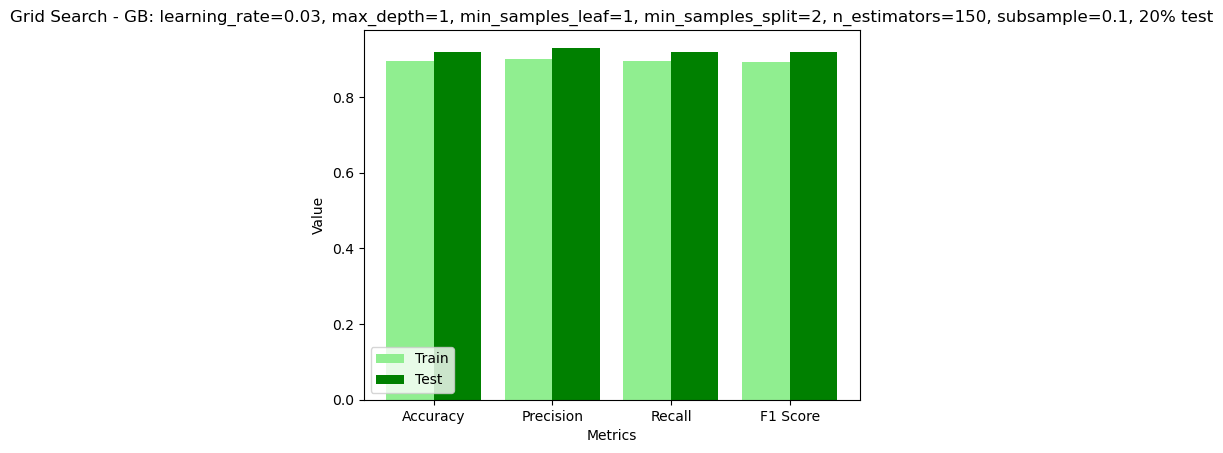
**Figure 8:** a) Metrics for k-NN (blue) and GB (purple), Tr: training set, Te: test set, Acc: accuracy, Prec: precision, Rec: recall, and F1: F1-score. b) Feature importance of GB.

The Grid Search for GB (figure 9) resulted in better accuracy in test than in the training set, which is counterintuitive for overfitting and the same happened with other metrics, thus the model may made right prediction 83% of the time. The parameter ‘subsample’ probably helped to reduce overfitting and improved the generalization because it introduces randomness to fit each tree. Inclusively, the ‘CO2 uptake’ became a feature of more importance in this model and when using the SMOTE.



**Figure 9:** a) Grid Search - GB: learning\_rate=0.03, max\_depth=1, min\_samples\_leaf=1, min\_samples\_split=2, n\_estimators=150, subsample=0.1, 20%. b) Feature importance Grid Search – GB.

Figure 10 considers the SMOTE and the class balance provided good performance in training and test sets with small difference, suggesting that is not overfitting. Consistent and slightly higher performance in all metrics on the test compared to the training suggest that the model can generalize well to new data. 93% of the samples predicted as positive were actually positive (precision), and 92% of the actual positive samples were predicted as positive (recall), indicating that the model performed well on the positive class, which is essential in this study as the objective is to identify which material can be suitable given a certain concentration of CO2 uptake desired by the industry.



**Figure 10:** a) Metrics. b) Confusion matrix. Condition: learning\_rate=0.03, max\_depth=1, min\_samples\_leaf=1, min\_samples\_split=2, n\_estimators=150, subsample=0.1, 20% test.

# **Conclusions**

The GB is robust and performed better than k-NN, and their performance was validated using cross-validation to ensure that the chosen hyperparameters generalized well to new data. Although the Grid Search with SMOTE helped to reduce the probability of overfitting and improved the generalization, there is still room for more progress. As the objective is to identify which material can be suitable given a certain concentration of CO2 uptake desired by the industry, improving the reliability of the positive class performance will be necessary, and increasing the number of records might solve it.

# **Future recommendations**

I recommend increasing the number of records per material to make the data more robust and increase the reliability of the results. And I suggest that the objective changes to answer which material would be suitable to adsorb a certain amount of CO2 instead of predicting the adsorption of CO2 in different materials, as all the materials analysed already have this capacity. In addition, it would be interesting if a company asked if I want to adsorb ‘X’ amount of CO2, what material should I use? Answering this question should help structure the company's planning.

# **Project's timeline overview**

Overall, the development of CA2 helped achieve many project milestones, particularly in gathering data, performing statistics, and implementing ML (Appendix 1). This process was satisfactory because my colleague and I will discuss our findings to reach the best option and finally deploy the model.

During the implementation phase, I faced many challenges, such as finding data, getting good metrics scores, and improving the models. But this helped me see that before proceeding with the following steps, I will have to increase the number of records to give the model more credibility.

# **Ethical considerations**

We will not intend to work with data that involves sensitive data, user privacy or potential social impacts; our data will be essential from laboratory research. The data will not be anonymised because they are from researchers who have done all the work and deserve to be appropriately recognised. In Appendix 2, we are presenting our Ethics Form signed.

# **Appendices**

Appendix 1:

Milestones achieved.

A diagram of a process

Description automatically generated

Appendix 2:

A paper with text and a checklist

Description automatically generated with medium confidence

# **References**

Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, [online] 16(16), pp.321–357. doi:https://doi.org/10.1613/jair.953.

Chawla, N.V. (2010). Data Mining for Imbalanced Datasets: An Overview. In: *Data Mining and Knowledge Discovery Handbook*. [online] pp.875–886. doi:https://doi.org/10.1007/978-0-387-09823-4\_45. [Accessed 13 Dec. 2023].

Daneshvar, E., Wicker, R.J., Show, P.-L. and Bhatnagar, A. (2022). Biologically-mediated carbon capture and utilization by microalgae towards sustainable CO2 biofixation and biomass valorization – A review. *Chemical Engineering Journal*, 427, p.130884. doi:https://doi.org/10.1016/j.cej.2021.130884.

Devore, J.L. (2012). *Probability and statistics for engineering and the sciences*. Belmont, Calif.] Brooks/Cole, Cengage Learning.

Dziejarski, B., Serafin, J., Andersson, K. and Krzyżyńska, R. (2023). CO2 capture materials: a review of current trends and future challenges. *Materials Today Sustainability*, 24, p.100483. doi:https://doi.org/10.1016/j.mtsust.2023.100483.

Furao Ren and Weijun Liu (2023). Review of CO2 Adsorption Materials and Utilization Technology. *Catalysts*, *13*(8), 1176. <https://doi.org/10.3390/catal13081176>

Muhammad Asif *et al.* (2018). Post-combustion CO2 capture with chemical absorption and hybrid system: current status and challenges. In *Greenhouse Gases: Science and Technology* (Vol. 8, Issue 6, pp. 998–1031). John Wiley and Sons Inc. <https://doi.org/10.1002/ghg.1823>

Müller, A. C. and Guido, S. (2017). *Introduction to machine learning with Python: a guide for data scientists*. 1st ed. United States of America. O’reilly Media.

scikit-learn. (n.d.). sklearn.preprocessing.RobustScaler. [online] Available at: <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.RobustScaler.html#sklearn.preprocessing.RobustScaler> [Accessed 2 Dec. 2023].

Thaddeus Huetteman, Michelle Bowman and Nancy Slater-Thompson (2016). Electricity. In *International Energy Outlook 2016 With Projections to 2040* (pp. 81–100). Energy Information Administration (EIA).

Yuan, X., Suvarna, M., Low, S., Dissanayake, P.D., Lee, K.B., Li, J., Wang, X. and Ok, Y.S. (2021). Applied Machine Learning for Prediction of CO2 Adsorption on Biomass Waste-Derived Porous Carbons. *Environmental Science & Technology*, 55(17), pp.11925–11936. doi:https://doi.org/10.1021/acs.est.1c01849.

Prusty, S., Patnaik, S. and Dash, S.K. (2022). SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer. *Frontiers in Nanotechnology*, 4. doi:https://doi.org/10.3389/fnano.2022.972421.

# **Supplementary information**

**Table S1:** Data of materials used for CO2 adsorption.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Material | Precursor | Conjugated microporous polymer | Activation agent | Activation temperature (°c) | BET surface area (m2/g) | Total pore volume (cm3/g) | CO2 uptake (mmol/g - 25°C) | Adsorption pressure bar | Reference |
| Carbon based | Pomegranate peels |  | KOH | 700 | 585 | 0.28 | 4.11 | 1 | (Serafin et al., 2017) |
| Carbon based | Carrot peels |  | KOH | 700 | 1379 | 0.58 | 4.18 | 1 | (Serafin et al., 2017) |
| Carbon based | Fern leaves |  | KOH | 700 | 1593 | 0.74 | 4.12 | 1 | (Serafin et al., 2017) |
| Carbon based | Black locust |  | KOH | 830 | 2064 | 0.98 | 3.75 | 1 | (Zhang et al., 2016) |
| Carbon based | Rice husk char |  | KOH | 780 | 2965 | 1.14 | 3.71 | 1 | (Li et al., 2015a) |
| Carbon based | Rice husk char |  | KOH | 640 | 774 | 0.41 | 3.53 | 1 | (Li et al., 2015a) |
| Carbon based | Rice husk char |  | KOH | 710 | 1041 | 0.53 | 4.16 | 1 | (Li et al., 2015a) |
| Carbon based | Rice husk char |  | KOH | 780 | 1199 | 0.48 | 3.87 | 1 | (Li et al., 2015a) |
| Carbon based | Longan shell |  | KOH | 800 | 3139 | 2.5 | 3.4 | 1 | (Wei et al., 2017) |
| Carbon based | Paulownia sawdust |  | KOH | 700 | 1643 | 0.857 | 5.8 | 1 | (Zhu et al., 2014) |
| Carbon based | Argan hard shells |  | NaOH | 850 | 1827 | 0.96 | 3.73 | 1 | (Boujibar et al., 2018) |
| Carbon based | Argan hard shells |  | KOH | 850 | 2251 | 1.04 | 5.51 | 1 | (Boujibar et al., 2018) |
| Carbon based | Wooden chopstick |  | KOH | 700 |  |  | 2.63 | 1 | (Phadungbut et al., 2022) |
| Carbon based | Polypodium vulgare |  | KOH | 800 | 1994 | 0.998 | 5.67 | 1 | (Serafin et al., 2017) |
| Carbon based | Common oak leader |  | KOH | 700 | 1842 | 0.91 | 5.67 | 1 | (Serafin and Cruz, 2022) |
| Carbon based | Walnut shell |  | KOH | 800 | 1868 | 1.06 | 5.17 | 1 | (Serafin et al., 2023) |
| Carbon based | Palm date seeds |  | KOH | 900 | 1906 | 1.06 | 5.44 | 1 | (Alazmi et al., 2021) |
| Carbon based | Palm date seeds |  | KOH | 900 | 2335 | 1.54 | 4.67 | 1 | (Alazmi et al., 2021) |
| Carbon based | Palm date seeds |  | H3PO4 | 900 | 1439 | 0.6 | 4.4 | 1 | (Alazmi et al., 2021) |
| Carbon based | Palm date seeds |  | H3PO5 | 900 | 1218 | 0.5 | 4 | 1 | (Alazmi et al., 2021) |
| Carbon based | Spent coffee ground |  | K2CO3 | 600 | 645 | 0.26 | 3.45 | 1 | (Kim et al., 2020) |
| Carbon based | Spent coffee ground |  | K2CO3 | 600 | 740 | 0.3 | 3.65 | 1 | (Kim et al., 2020) |
| Carbon based | Spent coffee ground |  | K2CO3 | 700 | 1259 | 0.52 | 4.33 | 1 | (Kim et al., 2020) |
| Carbon based | Spent coffee ground |  | K2CO3 | 700 | 1476 | 0.61 | 4.54 | 1 | (Kim et al., 2020) |
| Carbon based | Spent coffee ground |  | K2CO3 | 800 | 1692 | 0.71 | 4.46 | 1 | (Kim et al., 2020) |
| Carbon based | Spent coffee ground |  | K2CO3 | 800 | 2337 | 1.15 | 3.78 | 1 | (Kim et al., 2020) |
| Carbon based | Pine sawdust |  | KOH | 700 | 1728.66 | 0.7 | 4.21 | 1 | (Gao, Su and Gao, 2020) |
| Carbon based | Pine sawdust |  | KOH | 800 | 2279.52 | 0.99 | 3.46 | 1 | (Gao, Su and Gao, 2020) |
| Carbon based | Pine sawdust |  | KOH | 900 | 2330.89 | 1.91 | 2.45 | 1 | (Gao, Su and Gao, 2020) |
| Carbon based | Date |  | KOH | 800 | 2112 | 0.94 | 4.18 | 1 | (Li et al., 2019) |
| Carbon based | Date |  | KOH | 800 | 3255 | 1.65 | 3.35 | 1 | (Li et al., 2019) |
| Carbon based | Date |  | KOH | 800 | 3337 | 2.05 | 2.9 | 1 | (Li et al., 2019) |
| Carbon based | Date |  | KOH | 800 | 1634 | 0.76 | 4.14 | 1 | (Li et al., 2019) |
| Carbon based | Date |  | KOH | 800 | 2367 | 1.15 | 4.36 | 1 | (Li et al., 2019) |
| Carbon based | Date |  | KOH | 800 | 2844 | 1.63 | 3.65 | 1 | (Li et al., 2019) |
| Carbon based | Garlic peel |  | KOH | 700 | 1049 | 0.69 | 3.8 | 1 | (Huang et al., 2019) |
| Carbon based | Garlic peel |  | KOH | 700 | 1248 | 0.68 | 4.1 | 1 | (Huang et al., 2019) |
| Carbon based | Banana stems |  |  | 700 | 909 | 0.44 | 3.2 | 1 | (Sivadas, Damodaran and Raghavan, 2019) |
| Carbon based | Banana fiber |  |  | 500 | 1260 | 0.81 | 5 | 1 | (Sivadas, Damodaran and Raghavan, 2019) |
| Carbon based | Glucose |  | KOH | 500 | 972 | 0.49 | 3.01 | 1 | (Sun, Yang and Li, 2019) |
| Carbon based | Glucose |  | KOH | 600 | 1515 | 0.9 | 4.19 | 1 | (Sun, Yang and Li, 2019) |
| Carbon based | Glucose |  | KOH | 700 | 1815 | 1.02 | 3.91 | 1 | (Sun, Yang and Li, 2019) |
| Carbon based | Glucose |  | KOH | 800 | 2305 | 1.12 | 3.96 | 1 | (Sun, Yang and Li, 2019) |
| Carbon based | Peanut shell |  | KOH | 680 | 1713 | 0.73 | 4.41 | 1 | (Li et al., 2015b) |
| Carbon based | Peanut shell |  | KOH | 730 | 1893 | 0.79 | 4.22 | 1 | (Li et al., 2015b) |
| Carbon based | Peanut shell |  | KOH | 780 | 1871 | 0.8 | 3.92 | 1 | (Li et al., 2015b) |
| Carbon based | Sawdust |  | KOH | 600 | 1511 | 0.65 | 4.3 | 1 | (Hirst, Taylor and Mokaya, 2018) |
| Carbon based | Sawdust |  | KOH | 700 | 1830 | 0.78 | 4.9 | 1 | (Hirst, Taylor and Mokaya, 2018) |
| Carbon based | Sawdust |  | KOH | 800 | 2163 | 0.93 | 4.7 | 1 | (Hirst, Taylor and Mokaya, 2018) |
| Carbon based | Sawdust |  | KOH | 800 | 2610 | 1.15 | 4 | 1 | (Hirst, Taylor and Mokaya, 2018) |
| Carbon based | 40% food waste + 60% wood |  | KOH | 850 | 841.3 | 0.36 | 3.23 | 1 | (Igalavithana et al., 2020) |
| Carbon based | 40% food waste + 60% wood |  | KOH | 850 | 667.4 | 0.29 | 2.73 | 1 | (Igalavithana et al., 2020) |
| Metal-organic frameworks (MOFs) | Flexible microporous MOF |  |  |  |  |  | 1.11 | 1 | (Wu et al., 2010) |
| Metal-organic frameworks (MOFs) | 2D MOF |  |  |  | 340.8 |  | 1.9 | 1 | (Yan et al., 2012) |
| Metal-organic frameworks (MOFs) | Hydrated 6.5 wt%-Mg-MOF-74 |  |  |  |  |  | 6.18 | 1 | (Yu and Balbuena, 2013) |
| Metal-organic frameworks (MOFs) | Hydrated 13 wt%-Mg-MOF-74 |  |  |  |  |  | 4.73 | 1 | (Yu and Balbuena, 2013) |
| Metal-organic frameworks (MOFs) | SNU-1100 |  |  |  | 411 | 0.248 | 0.58 | 1 | (Hong and Suh, 2012) |
| Metal-organic frameworks (MOFs) | Noninterpenetrated-SNU-700' |  |  |  | 5290 | 2.17 | 0.8 | 1 | (Prasad and Suh, 2012) |
| Metal-organic frameworks (MOFs) | Interpenetrated-SNU-710' |  |  |  | 1770 | 0.709 | 1.05 | 1 | (Prasad and Suh, 2012) |
| Metal-organic frameworks (MOFs) | ZNJU-43a |  |  |  | 2243 | 0.8943 | 4.6 | 1 | (Song et al., 2015) |
| Metal-organic frameworks (MOFs) | UPC-105 |  |  |  | 2082 |  | 2.37 | 1 | (Fan et al., 2018) |
| Metal-organic frameworks (MOFs) | UPC-106 |  |  |  | 1984 |  | 2.42 | 1 | (Fan et al., 2018) |
| Metal-organic frameworks (MOFs) | UPC-107 |  |  |  | 1865 |  | 2.06 | 1 | (Fan et al., 2018) |
| Metal-organic frameworks (MOFs) | UPC-108 |  |  |  | 1837 |  | 2.04 | 1 | (Fan et al., 2018) |
| Metal-organic frameworks (MOFs) | UPC-109 |  |  |  | 1601 |  | 1.08 | 1 | (Fan et al., 2018) |
| Metal-organic frameworks (MOFs) | UPC-110 |  |  |  | 1384 |  | 1.08 | 1 | (Fan et al., 2018) |
| Metal-organic frameworks (MOFs) | UPC-111 |  |  |  | 1732 |  | 1.88 | 1 | (Fan et al., 2018) |
| Metal-organic frameworks (MOFs) | UPC-112 |  |  |  | 1559 |  | 1.83 | 1 | (Fan et al., 2018) |
| Metal-organic frameworks (MOFs) | pt-UiO-66(Zr)(OH)2 |  |  |  | 1230 |  | 5.63 | 1 | (Hu et al., 2017) |
| Metal-organic frameworks (MOFs) | ZNJU-19 |  |  |  | 2165 | 0.882 | 4.75 | 1 | (Xu et al., 2020b) |
| Metal-organic frameworks (MOFs) | ZNJU-20 |  |  |  | 2154 | 0.902 | 4.63 | 1 | (Xu et al., 2020b) |
| Metal-organic frameworks (MOFs) | SNU-77H |  |  |  | 3670 | 1.52 | 0.89 | 1 | (Park et al., 2011) |
| Metal-organic frameworks (MOFs) | HCM-Cu3(BTC)2/Hierarchical porous carbon monoliths |  |  |  | 516 | 0.26 | 2.75 | 1 | (Qian et al., 2012) |
| Microporous polymers |  | CMP@1 |  |  | 346 | 0.22 | 1.03 | 1 | (Xu et al., 2020a) |
| Microporous polymers |  | CMP@2 |  |  | 325 | 0.6 | 1.47 | 1 | (Xu et al., 2020a) |
| Microporous polymers |  | CMP@3 |  |  | 343 | 0.7 | 0.71 | 1 | (Xu et al., 2020a) |
| Microporous polymers |  | CMP |  |  | 772 | 1.21 | 1.61 | 1 | (Xie et al., 2013) |
| Microporous polymers |  | Co-CMP |  |  | 965 | 2.81 | 1.8 | 1 | (Xie et al., 2013) |
| Microporous polymers |  | Al-CMP |  |  | 798 | 1.41 | 1.74 | 1 | (Xie et al., 2013) |
| Microporous polymers |  | ZnPe50%F-CMPs |  |  | 240 | 0.376 | 2.05 | 1 | (Cui, Yao and Xu, 2017) |
| Microporous polymers |  | CMP-1 |  |  | 837 | 0.45 | 1.18 | 1 | (Dawson, Adams and Cooper, 2011) |
| Microporous polymers |  | CMP-1 COOH |  |  | 522 | 0.3 | 1.6 | 1 | (Dawson, Adams and Cooper, 2011) |
| Microporous polymers |  | CMP-1-(CH3)2 |  |  | 899 | 0.75 | 0.94 | 1 | (Dawson, Adams and Cooper, 2011) |
| Microporous polymers |  | CMP-1-(OH)2 |  |  | 1043 | 0.71 | 1.07 | 1 | (Dawson, Adams and Cooper, 2011) |
| Microporous polymers |  | NCMP-2 |  |  | 900 | 0.55 | 1.15 | 1 | (Jiang et al., 2009) |
| Microporous polymers |  | TCMP-0 |  |  | 963 | 0.98 | 1.34 | 1 | (Ren et al., 2012) |
| Microporous polymers |  | TNCMP-2 |  |  | 995 | 0.55 | 1.45 | 1 | (Ren et al., 2012) |
| Microporous polymers |  | TCMP-3 |  |  | 961 | 0.36 | 1.26 | 1 | (Ren et al., 2012) |
| Microporous polymers |  | TCMP-5 |  |  | 494 | 0.51 | 1.22 | 1 | (Ren et al., 2012) |
| Microporous polymers |  | CMP-0 |  |  | 1018 | 0.56 | 1.21 | 1 | (Jiang et al., 2008) |
| Microporous polymers |  | CMP-3 |  |  | 522 | 0.26 | 0.68 | 1 | (Jiang et al., 2008) |
| Microporous polymers |  | CMP-5 |  |  | 512 | 0.47 | 0.68 | 1 | (Jiang et al., 2008) |

# **References for supplementary information**

Alazmi, A., Nicolae, S.A., Modugno, P., Hasanov, B.E., Titirici, M.M. and Costa, P.M.F.J. (2021). Activated Carbon from Palm Date Seeds for CO2 Capture. *International Journal of Environmental Research and Public Health*, 18(22), p.12142. doi:https://doi.org/10.3390/ijerph182212142.

Boujibar, O., Souikny, A., Ghamouss, F., Achak, O., Dahbi, M. and Chafik, T. (2018). CO2 capture using N-containing nanoporous activated carbon obtained from argan fruit shells. *Journal of Environmental Chemical Engineering*, 6(2), pp.1995–2002. doi:https://doi.org/10.1016/j.jece.2018.03.005.

Cui, D., Yao, C. and Xu, Y.-H. (2017). Conjugated microporous polymers with azide groups: a new strategy for postsynthetic fluoride functionalization and effectively enhanced CO2 adsorption properties. *Chemical Communications*, 53(83), pp.11422–11425. doi:https://doi.org/10.1039/c7cc06528k.

Dawson, R., Adams, D.J. and Cooper, A.I. (2011). Chemical tuning of CO2 sorption in robust nanoporous organic polymers. *Chemical Science*, 2(6), p.1173. doi:https://doi.org/10.1039/c1sc00100k.

Fan, W., Wang, X., Liu, X., Ben Bin Xu, Zhang, X., Wang, W., Wang, Y., Dai, F., Yuan, D. and Sun, D. (2018). Regulating C2H2 and CO2 Storage and Separation through Pore Environment Modification in a Microporous Ni-MOF. *ACS Sustainable Chemistry & Engineering*, 7(2), pp.2134–2140. doi:https://doi.org/10.1021/acssuschemeng.8b04783.

Gao, N., Su, R. and Gao, N. (2020). Preparation of activated biomass carbon from pine sawdust for supercapacitor and CO2 capture. *International Journal of Energy Research*, 44(6), pp.4335–4351. doi:https://doi.org/10.1002/er.5206.

Hirst, E.A., Taylor, A. and Mokaya, R. (2018). A simple flash carbonization route for conversion of biomass to porous carbons with high CO2 storage capacity. *Journal of Materials Chemistry A*, 6(26), pp.12393–12403. doi:https://doi.org/10.1039/c8ta04409k.

Hong, D.H. and Suh, M.P. (2012). Selective CO2 adsorption in a metal–organic framework constructed from an organic ligand with flexible joints. *Chemical Communications*, 48(73), p.9168. doi:https://doi.org/10.1039/c2cc34482c.

Hu, Z., Wang, Y., Farooq, S. and Zhao, D. (2017). A highly stable metal‐organic framework with optimum aperture size for CO2 capture. *AIChE Journal*, 63(9), pp.4103–4114. doi:https://doi.org/10.1002/aic.15837.

Huang, G., Wu, X., Hou, Y. and Cai, J. (2019). Sustainable porous carbons from garlic peel biowaste and KOH activation with an excellent CO2 adsorption performance. *Biomass conversion and biorefinery*, 10(2), pp.267–276. doi:https://doi.org/10.1007/s13399-019-00412-6.

Igalavithana, A.D., Choi, S.W., Dissanayake, P.D., Shang, J., Wang, C.-H., Yang, X., Kim, S., Tsang, D.C.W., Lee, K.B. and Ok, Y.S. (2020). Gasification biochar from biowaste (food waste and wood waste) for effective CO2 adsorption. *Journal of Hazardous Materials*, 391, pp.121147–121147. doi:https://doi.org/10.1016/j.jhazmat.2019.121147.

Jiang, J.-X., Su, F., Trewin, A., Wood, C.D., Niu, H., Jones, J.T.A., Khimyak, Y.Z. and Cooper, A.I. (2008). Synthetic Control of the Pore Dimension and Surface Area in Conjugated Microporous Polymer and Copolymer Networks. *Journal of the American Chemical Society*, 130(24), pp.7710–7720. doi:https://doi.org/10.1021/ja8010176.

Jiang, J.-X., Trewin, A., Su, F., Wood, C.D., Niu, H., Jones, J.T.A., Khimyak, Y.Z. and Cooper, A.I. (2009). Microporous Poly(tri(4-ethynylphenyl)amine) Networks: Synthesis, Properties, and Atomistic Simulation. *Macromolecules*, 42(7), pp.2658–2666. doi:https://doi.org/10.1021/ma802625d.

Kim, M.-J., Choi, S.W., Kim, H., Mun, S. and Lee, K.B. (2020). Simple synthesis of spent coffee ground-based microporous carbons using K2CO3 as an activation agent and their application to CO2 capture. *Chemical Engineering Journal*, [online] 397, p.125404. doi:https://doi.org/10.1016/j.cej.2020.125404.

Li, D., Ma, T., Zhang, R., Tian, Y. and Qiao, Y. (2015a). Preparation of porous carbons with high low-pressure CO2 uptake by KOH activation of rice husk char. *Fuel*, [online] 139, pp.68–70. doi:https://doi.org/10.1016/j.fuel.2014.08.027.

Li, D., Tian, Y., Li, L., Li, J. and Zhang, H. (2015b). Production of highly microporous carbons with large CO2 uptakes at atmospheric pressure by KOH activation of peanut shell char. *Journal of Porous Materials*, 22(6), pp.1581–1588. doi:https://doi.org/10.1007/s10934-015-0041-7.

Li, J., Michalkiewicz, B., Min, J., Ma, C., Chen, X., Gong, J., Mijowska, E. and Tang, T. (2019). Selective preparation of biomass-derived porous carbon with controllable pore sizes toward highly efficient CO2 capture. *Chemical Engineering Journal*, 360, pp.250–259. doi:https://doi.org/10.1016/j.cej.2018.11.204.

Park, H.J., Lim, D.-W., Yang, W.S., Oh, T.-R. and Suh, M.P. (2011). A Highly Porous Metal–Organic Framework: Structural Transformations of a Guest‐Free MOF Depending on Activation Method and Temperature. *Chemistry*, 17(26), pp.7251–7260. doi:https://doi.org/10.1002/chem.201003376.

Phadungbut, P., Koo-amornpattana, W., Bumroongsri, P., Ratchahat, S., Kunthakudee, N., Jonglertjunya, W., Chalermsinsuwan, B. and Hunsom, M. (2022). Adsorptive purification of CO2/H2 gas mixtures of spent disposable wooden chopstick-derived activated carbon: Optimal synthesis condition. *Separation and Purification Technology*, 291, p.120948. doi:https://doi.org/10.1016/j.seppur.2022.120948.

Prasad, T.K. and Suh, M.P. (2012). Control of Interpenetration and Gas-Sorption Properties of Metal-Organic Frameworks by a Simple Change in Ligand Design. *Chemistry - A European Journal*, 18(28), pp.8673–8680. doi:https://doi.org/10.1002/chem.201200456.

Qian, D., Lei, C., Hao, G.-P., Li, W.-C. and Lu, A.-H. (2012). Synthesis of Hierarchical Porous Carbon Monoliths with Incorporated Metal–Organic Frameworks for Enhancing Volumetric Based CO2 Capture Capability. *ACS Applied Materials & Interfaces*, 4(11), pp.6125–6132. doi:https://doi.org/10.1021/am301772k.

Ren, S., Ogden, M.D., Laybourn, A., Jiang, J.-X., Khimyak, Y.Z., Adams, D.J. and Cooper, A.I. (2012). Functional conjugated microporous polymers: from 1,3,5-benzene to 1,3,5-triazine. *Polymer Chemistry*, 3(4), pp.928–928. doi:https://doi.org/10.1039/c2py00585a.

Serafin, J. and Cruz, O.F. (2022). Promising activated carbons derived from common oak leaves and their application in CO2 storage. *Journal of environmental chemical engineering*, 10(3), pp.107642–107642. doi:https://doi.org/10.1016/j.jece.2022.107642.

Serafin, J., Dziejarski, B., Cruz Junior, O.F. and Sreńscek-Nazzal, J. (2023). Design of highly microporous activated carbons based on walnut shell biomass for H2 and CO2 storage. *Carbon*, 201, pp.633–647. doi:https://doi.org/10.1016/j.carbon.2022.09.013.

Serafin, J., Kiełbasa, K. and Michalkiewicz, B. (2022). The new tailored nanoporous carbons from the common polypody (Polypodium vulgare): The role of textural properties for enhanced CO2 adsorption. *Chemical Engineering Journal*, 429, p.131751. doi:https://doi.org/10.1016/j.cej.2021.131751.

Serafin, J., Narkiewicz, U., Morawski, A.W., Wróbel, R.J. and Michalkiewicz, B. (2017). Highly microporous activated carbons from biomass for CO2 capture and effective micropores at different conditions. *Journal of CO2 Utilization*, 18, pp.73–79. doi:https://doi.org/10.1016/j.jcou.2017.01.006.

Sivadas, D.L., Damodaran, A. and Raghavan, R. (2019). Microporous Carbon Monolith and Fibre from Freeze Dried Banana Stem for High Efficiency Carbon Dioxide Adsorption. *ACS Sustainable Chemistry & Engineering*, 7(15), pp.12807–12816. doi:https://doi.org/10.1021/acssuschemeng.9b01653.

Song, C., Hu, J., Ling, Y., Feng, Y., Krishna, R., Chen, D. and He, Y. (2015). The accessibility of nitrogen sites makes a difference in selective CO2 adsorption of a family of isostructural metal–organic frameworks. *Journal of Materials Chemistry A*, 3(38), pp.19417–19426. doi:https://doi.org/10.1039/c5ta05481h.

Sun, H., Yang, B. and Li, A. (2019). Biomass derived porous carbon for efficient capture of carbon dioxide, organic contaminants and volatile iodine with exceptionally high uptake. *Chemical Engineering Journal*, 372, pp.65–73. doi:https://doi.org/10.1016/j.cej.2019.04.061.

Wei, H., Chen, H., Fu, N., Chen, J., Lan, G., Qian, W., Liu, Y., Lin, H. and Han, S. (2017). Excellent electrochemical properties and large CO2 capture of nitrogen-doped activated porous carbon synthesised from waste longan shells. *Electrochimica Acta*, 231, pp.403–411. doi:https://doi.org/10.1016/j.electacta.2017.01.194.

Wu, H., Reali, R.S., Smith, D.A. and Trachtenberg, M.C. (2010). Highly Selective CO2 Capture by a Flexible Microporous Metal–Organic Framework (MMOF) Material. *Chemistry European Journal*, 16(47), pp.13951–13954. doi:https://doi.org/10.1002/chem.201002683.

Xie, Y., Wang, T.-T., Liu, X.-H., Zou, K. and Deng, W.-Q. (2013). Capture and conversion of CO2 at ambient conditions by a conjugated microporous polymer. *Nature Communications*, 4(1). doi:https://doi.org/10.1038/ncomms2960.

Xu, C., Zhu, Y., Yao, C., Xie, W., Xu, G., Zhang, S., Zhao, Y. and Xu, Y. (2020a). Facile synthesis of tetraphenylethene-based conjugated microporous polymers as adsorbents for CO2 and organic vapor uptake. *New journal of chemistry*, 44(2), pp.317–321. doi:https://doi.org/10.1039/c9nj04562g.

Xu, T., Fan, L., Jiang, Z., Zhou, P., Li, Z., Lu, H. and He, Y. (2020b). Immobilization of N-oxide functionality into NbO-type MOFs for significantly enhanced C2H2/CH4 and CO2/CH4 separations. *Dalton Transactions*, [online] 49(21), pp.7174–7181. doi:https://doi.org/10.1039/D0DT01081B.

Yan, Q., Lin, Y., Wu, P., Zhao, L., Cao, L., Peng, L., Kong, C. and Chen, L. (2012). Designed Synthesis of Functionalized Two-Dimensional Metal-Organic Frameworks with Preferential CO2 Capture. *ChemPlusChem*, 78(1), pp.86–91. doi:https://doi.org/10.1002/cplu.201200270.

Yu, J. and Balbuena, P.B. (2013). Water Effects on Postcombustion CO2 Capture in Mg-MOF-74. *The Journal of Physical Chemistry C*, 117(7), pp.3383–3388. doi:https://doi.org/10.1021/jp311118x.

Zhang, C., Song, W., Ma, Q., Xie, L., Zhang, X. and Guo, H. (2016). Enhancement of CO2 Capture on Biomass-Based Carbon from Black Locust by KOH Activation and Ammonia Modification. *Energy & Fuels*, 30(5), pp.4181–4190. doi:https://doi.org/10.1021/acs.energyfuels.5b02764.

Zhu, X.-L., Wang, P.-Y., Peng, C., Yang, J. and Yan, X.-B. (2014). Activated carbon produced from paulownia sawdust for high-performance CO2 sorbents. *Chinese Chemical Letters*, 25(6), pp.929–932. doi:https://doi.org/10.1016/j.cclet.2014.03.039.