Prediction of CO_2 solubility in ionic liquids using machine learning methods

Name: Vijendra Singh Bhati

Roll no: 210107095

Course: CL653, Applications of AI & ML in Chemical Engineering

Date of Submission: 14/04/2024

1. Project Overview

Introduction:

In recent years, the utilization of ionic liquids (ILs) as environmentally benign solvents has gained significant attention across various industrial sectors. Ionic liquids possess unique properties such as low volatility, high thermal stability, and tunable chemical structures, making them promising candidates for numerous applications including gas capture, separation processes, and catalysis. Among these applications, the ability to predict the solubility of carbon dioxide (CO2) in ionic liquids has emerged as a crucial area of research due to its implications in carbon capture and storage (CCS), enhanced oil recovery (EOR), and green chemical synthesis.

The solubility of CO2 in ionic liquids plays a pivotal role in determining the efficiency and feasibility of CO2 capture technologies. Accurate prediction of CO2 solubility in ILs can aid in the design and optimization of processes aimed at reducing CO2 emissions and mitigating climate change. By developing robust AI and machine learning models capable of predicting CO2 solubility in ionic liquids, researchers and engineers can expedite the discovery and development of novel IL-based CO2 capture systems, thereby contributing to the transition towards a sustainable energy future.

Objectives of the Project:

The primary objectives of this project are as follows:

1. Develop an AI/ML model capable of accurately predicting the solubility of CO2 in ionic liquids under varying conditions such as temperature, pressure, and IL composition.
2. Investigate the underlying molecular interactions and structural features that influence CO2 solubility in ionic liquids, utilizing advanced data analytics techniques.
3. Validate the performance of the developed model using experimental data from literature and/or computational simulations, ensuring its reliability and applicability across different IL systems.
4. Explore the potential of the developed AI/ML model for guiding the design and optimization of IL-based CO2 capture processes, with a focus on enhancing efficiency and reducing energy consumption.

By achieving these objectives, this project aims to contribute to the advancement of CO2 capture technologies and facilitate the transition towards a more sustainable and environmentally conscious energy landscape.

2. Description of Project

Theoretical Background:

Ionic liquids (ILs) are a class of molten salts composed entirely of ions, typically consisting of bulky organic cations and various anions. Due to their unique combination of properties such as negligible vapor pressure, high thermal stability, and tunable solvation behavior, ILs have garnered widespread attention as green solvents in diverse applications ranging from catalysis to electrochemistry. One particularly promising area of research involves the use of ILs for carbon dioxide (CO2) capture and separation.

The solubility of CO2 in ILs is a critical parameter that determines the efficiency of IL-based CO2 capture processes. Understanding the factors that influence CO2 solubility in ILs is essential for designing and optimizing these processes. The solvation behavior of CO2 in ILs is governed by a complex interplay of molecular interactions, including electrostatic interactions, hydrogen bonding, and van der Waals forces. Predicting CO2 solubility accurately requires consideration of various factors such as temperature, pressure, IL composition, and structural properties of both CO2 and ILs.

Specific Problem Statement:

The specific problem addressed in this project is the development of an AI/ML model capable of predicting the solubility of CO2 in ionic liquids with high accuracy and reliability. This involves the creation of a predictive model that can capture the complex relationships between CO2 solubility and the multitude of factors that influence it, including temperature, pressure, IL composition, and structural features of CO2 and ILs.

Significance of Addressing this Issue:

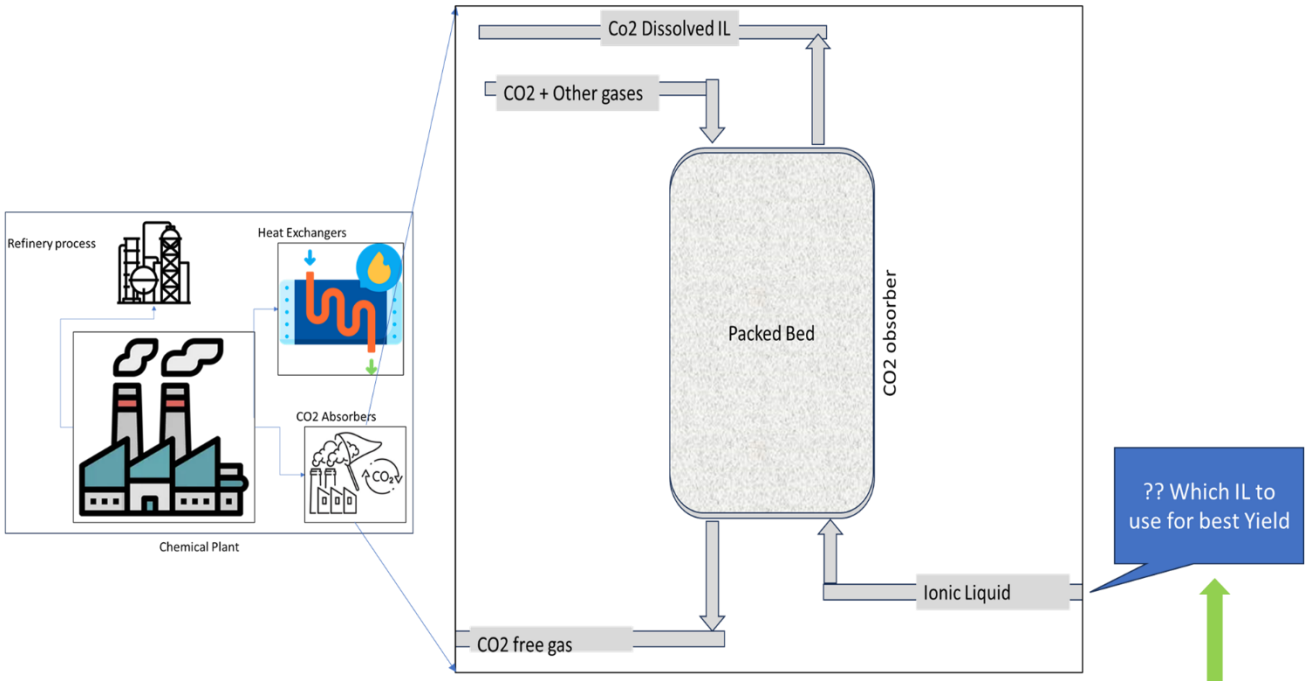
The significance of addressing the prediction of CO2 solubility in ILs lies in its potential impact on advancing CO2 capture technologies and mitigating climate change. By accurately predicting CO2 solubility in ILs, researchers and engineers can:

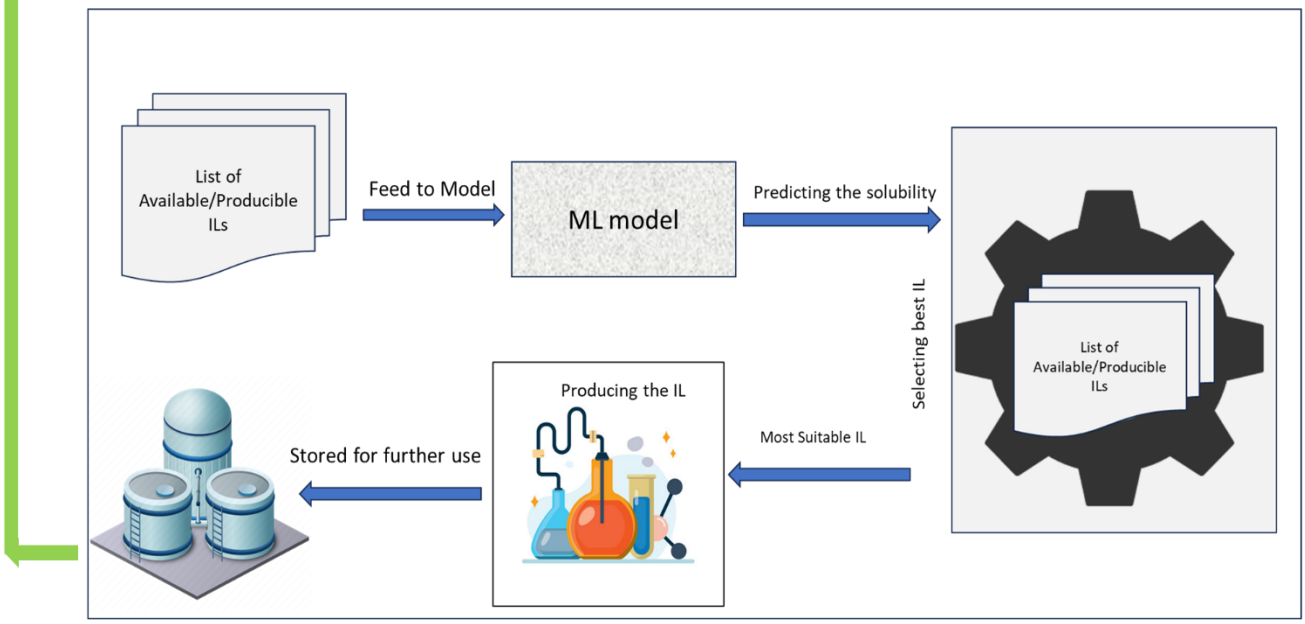
1. **Optimize CO2 Capture Processes:** The development of accurate predictive models enables the design and optimization of IL-based CO2 capture processes, leading to improved efficiency and reduced energy consumption.
2. **Facilitate Material Design:** Understanding the molecular interactions that govern CO2 solubility in ILs can guide the rational design of novel ILs with enhanced CO2 capture properties, thus accelerating the discovery of more efficient CO2 sorbents.
3. **Contribute to Sustainable Development:** IL-based CO2 capture technologies have the potential to play a significant role in reducing greenhouse gas emissions and mitigating climate change. By advancing our understanding of CO2 solubility in ILs, this project contributes to the development of more sustainable and environmentally friendly solutions for CO2 capture and utilization.

Overall, addressing the prediction of CO2 solubility in ILs is essential for advancing the field of CO2 capture and contributing to global efforts to combat climate change.

3. Block Diagram and Flow Charts

Chemical Plant Flowsheet





Selection of the IL to use

Model Development and Deployment

Data Collection

- DataBases
- Experimental Data

Data Preprocessing

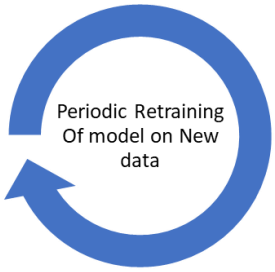
- Data Cleaning
- Normalisation
- Feature engineering

Model Development

- Feature Selection
- Model Training
- Model Evaluation

Model Deployment:

- Integration with CO2 Capture System:
- User Interface Development



Process Optimisation

Utilization of optimization algorithms to optimize CO2 capture process parameters based on model predictions.

Incorporation of model predictions into decision-making processes to enhance the efficiency and cost-effectiveness of CO2 capture operations.

Recommendations for optimizing process parameters to maximize CO2 capture efficiency.

4. Data Sources

For this project, the data sources for obtaining information on CO2 solubility in ionic liquids will primarily include:

1. **Literature Review:** Compilation of experimental data from published research articles, journals, and conference proceedings focusing on CO2 solubility in various ionic liquids.

   Zhigang Lei, Chenga Dai, Biaohua Chen: Gas Solubility in Ionic Liquid: November 6, 2013

   Prediction of CO2 solubility in ionic liquids using machine learning methods : Zhen Song , Huaiwei Shi , Xiang Zhang, Teng Zhou : 30 April 2020

2. **Online Databases:** Accessing publicly available databases or repositories containing experimental data on CO2 solubility in ionic liquids, such as National Institute of Standards and Technology (NIST) Standard Reference Database #147: https://ilthermo.boulder.nist.gov/

Data Characteristics:

Volume:

- The volume of data for CO2 solubility in ionic liquids may vary depending on the number of experimental studies available and the diversity of IL systems studied. It may range from a few hundred data points to several thousand.

Variety:

- The data comprises of a variety of attributes including:
    - Chemical properties of ionic liquids (e.g., cation/anion structure, molar mass, density).
    - Thermodynamic conditions (e.g., temperature, pressure) at which solubility measurements were conducted.
    - Structural features of CO2 molecules (e.g., size, shape) and their interaction with ILs.
    - Experimental techniques used for solubility measurement (e.g., gravimetric, spectroscopic).

Velocity:

- The velocity of data acquisition vary depending on the frequency of new experimental studies being published or updates to computational simulations. While the volume of data may not increase rapidly, the velocity of updates and additions to the dataset may be moderate, with new data becoming available periodically as new research is conducted or published.

Overall, the data for this project will encompass a diverse range of attributes related to CO2 solubility in ionic liquids, obtained from experimental measurements , with varying volumes, varieties, and velocities of data acquisition.

5. Description of Data:

Nature of Data: The data for predicting the solubility of CO2 in ionic liquids can be considered as steady-state. Steady-state data refers to observations or measurements taken when a system has reached a stable equilibrium under constant conditions. In the context of CO2 solubility in ionic liquids, this means that the measurements of CO2 solubility are typically taken at specific temperature and pressure conditions, where the system has reached a stable state with respect to CO2 dissolution in the IL.

Implications for the Project: The steady-state nature of the data implies that the relationships between CO2 solubility and various factors such as temperature, pressure, and IL composition can be analyzed without considering transient effects or time-dependent behavior. This simplifies the modeling process as it allows for the development of static predictive models based on steady-state data points.

Data Preprocessing: In preparing the data for analysis, several preprocessing steps may be necessary to ensure the quality and suitability of the data for model development. These steps include:

1. **Data Cleaning:**
    - Removing any outliers or erroneous data points that may skew the analysis.
    - Handling missing values by imputation or deletion, depending on the extent of missingness and the impact on the dataset.
2. **Normalization/Scaling:**
    - Scaling numerical features to a common range to prevent certain features from dominating the model training process. Temperatures and pressures Columns were scaled
3. **Feature Engineering:**
    - 51 new columns are added (features are engineered) to represent the ionic compound as numeric data to be fed into the ANN

	IL	cation	anion	x_CO2	T (K)	P (bar)	[CH3]	[CH2]	[CH]	[OCH2]	...	[MeSO3]	[TfO]	[NfO]	[TDfO]	[TOS]	[C12PhSO3]	[DMPO4]	[DEPO4]	[DBPO4]	[methide]
0	[BMIM][BF4]	[BMIM]	[BF4]	0.61	363.15	246.0	1.0	3.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	[BMIM][BF4]	[BMIM]	[BF4]	0.50	383.15	235.0	1.0	3.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	[BMIM][BF4]	[BMIM]	[BF4]	0.61	353.15	223.3	1.0	3.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	[BMIM][BF4]	[BMIM]	[BF4]	0.50	373.15	198.0	1.0	3.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	[BMIM][BF4]	[BMIM]	[BF4]	0.61	343.15	188.5	1.0	3.0	0.0	0.0	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

4. **Encoding Categorical Variables:**
    - Encoding categorical variables such as the identities of ionic liquid components into numerical representations suitable for model training, using techniques like one-hot encoding or label encoding.

6. Strategies for AI/ML Model Development

Model Selection: For predicting the solubility of CO2 in ionic liquids, several AI/ML models can be considered, including:

1. **Support Vector Regression (SVR):** SVR is effective for handling high-dimensional data and can capture complex relationships between input features and target variable.
2. **Neural Networks:** Deep learning models such as feedforward neural networks or recurrent neural networks (RNNs) can learn complex patterns from data, potentially capturing intricate relationships between CO2 solubility and various input features.

Training: For model training, we will use standard techniques such as stochastic gradient descent (SGD) or Adam optimization for neural networks, and built-in optimization algorithms for other models like Random Forest and SVR. We will utilize popular machine learning libraries such as Scikit-learn for neural networks.

Evaluation and Validation:

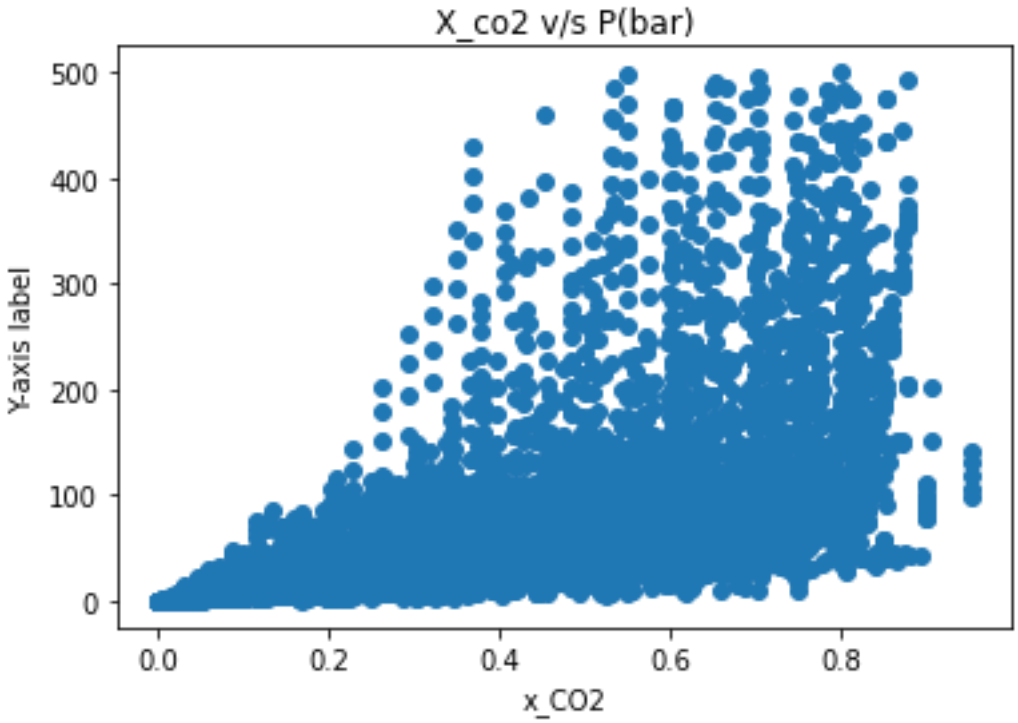
Evaluation Metrics: The following evaluation metrics will be used to assess the performance of the model:

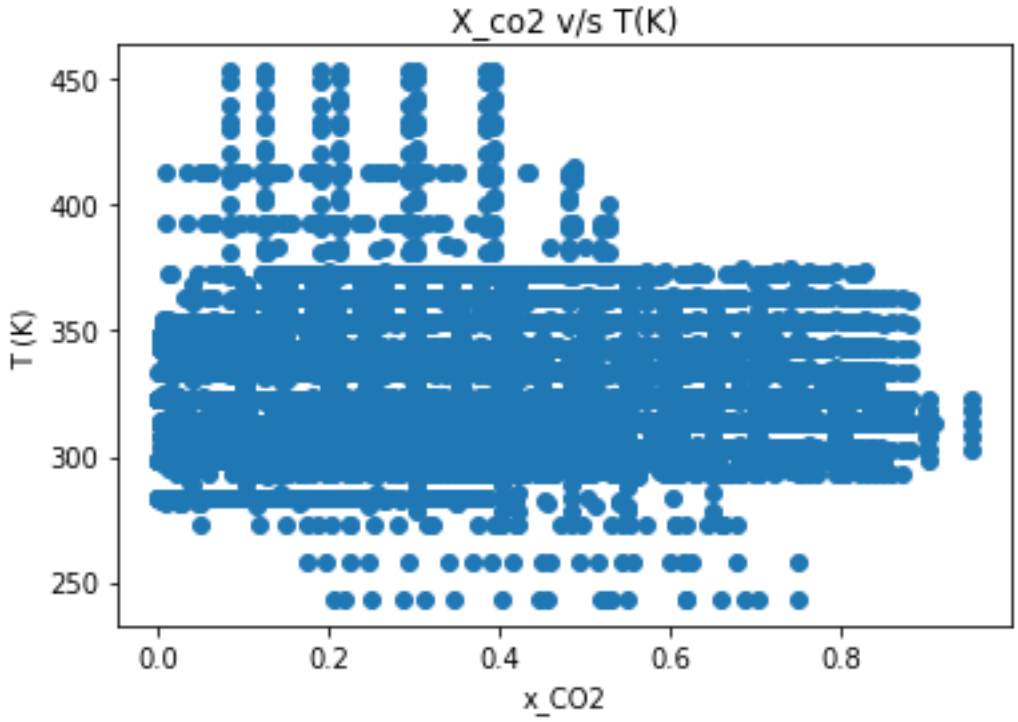
1. **Root Mean Squared Error (RMSE):** Measures the square root of the average of the squared differences between predicted and actual values, providing insight into the overall accuracy of the model.

Validation Strategy: To validate the model and ensure its generalizability and robustness, the following strategies will be employed:

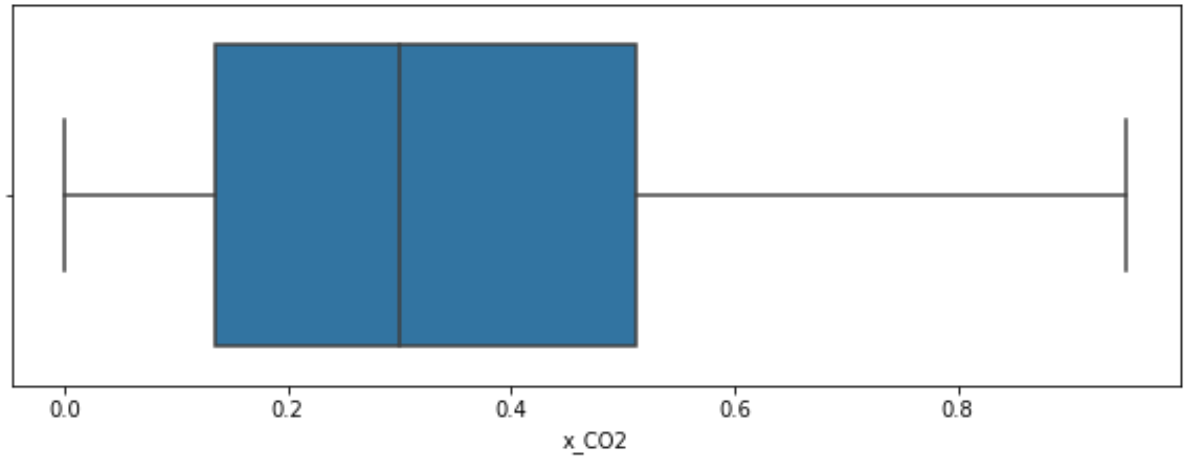
1. **Train-Test Split:** The dataset is randomly divided into training and testing sets, with typically 80% of the data used for training and the remaining 20% for testing.
2. **Cross-Validation:** K-fold cross-validation will be performed to assess the model's performance across multiple splits of the data. This helps in obtaining more reliable estimates of model performance and detecting overfitting.

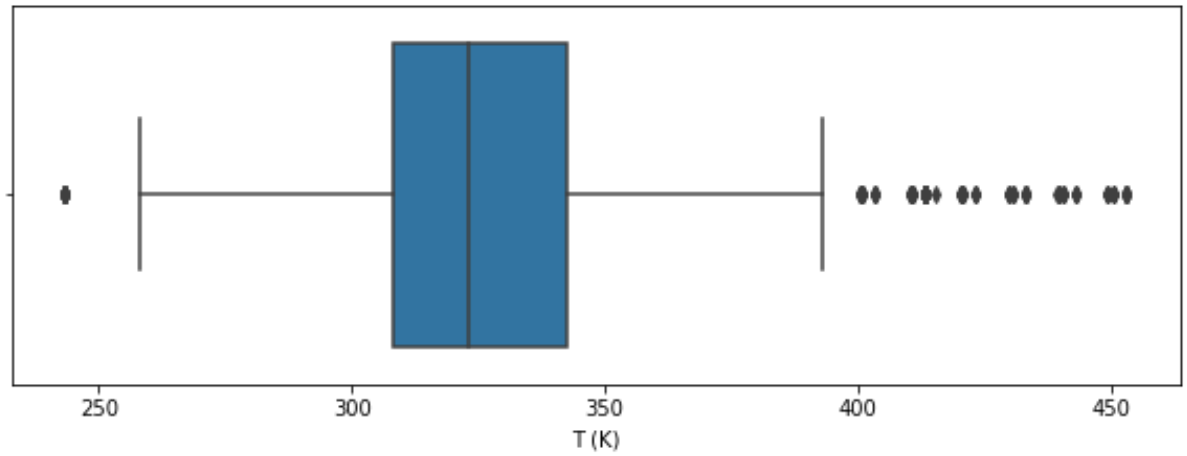
Basic Plots

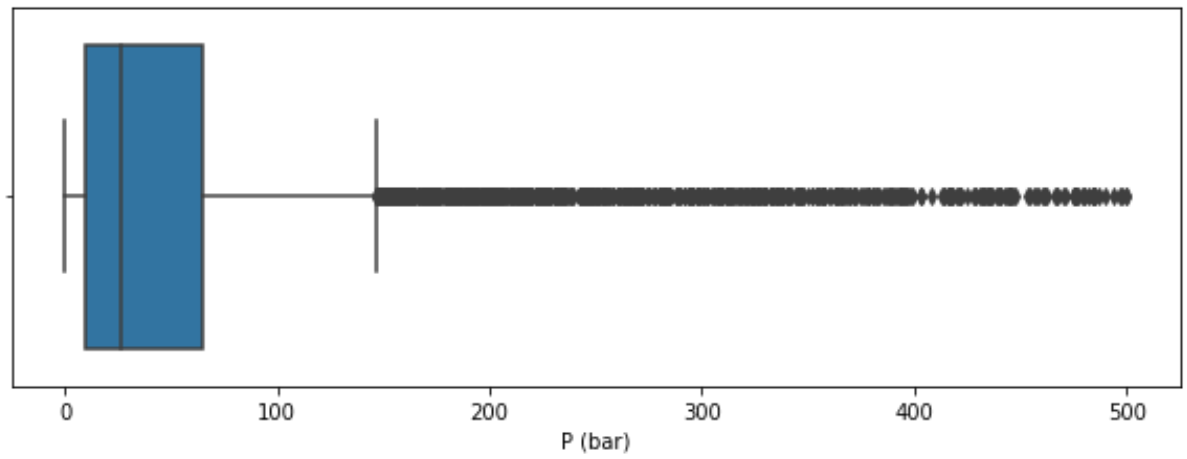

X_co2 v/s P(bar)
Y-axis label
500
400
300
200
100
0
0.0
0.2
0.4
0.6
0.8
x_CO2


X_co2 v/s T(K)
T (K)
450
400
350
300
250
0.0
0.2
0.4
0.6
0.8
x_CO2

Box Plots :


0.0
0.2
0.4
0.6
0.8
x_CO2


250
300
350
400
450
T (K)


0
100
200
300
400
500
P (bar)

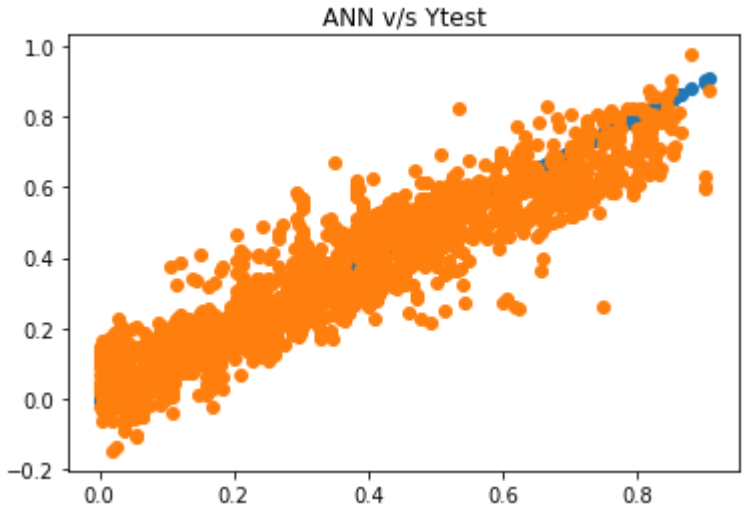
Results

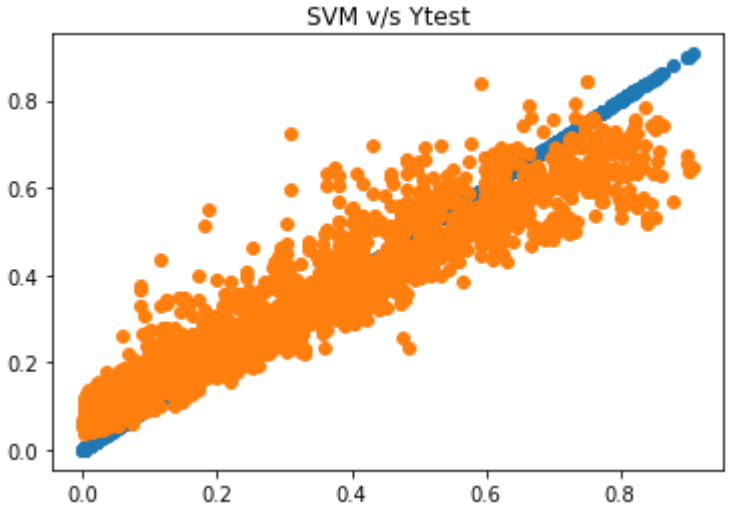
SVM performs better and is able to generalize better for different train test splits

Run:1

ANN RMSE: 0.1628774657156774

SVM RMSE: 0.08664705044469487

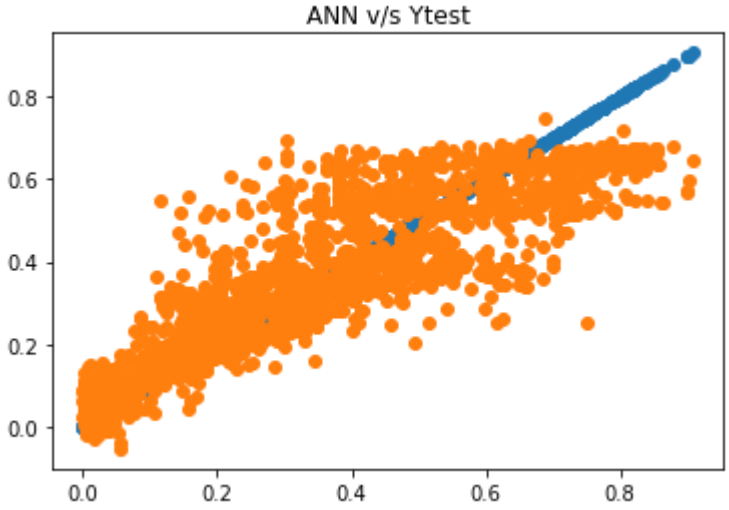


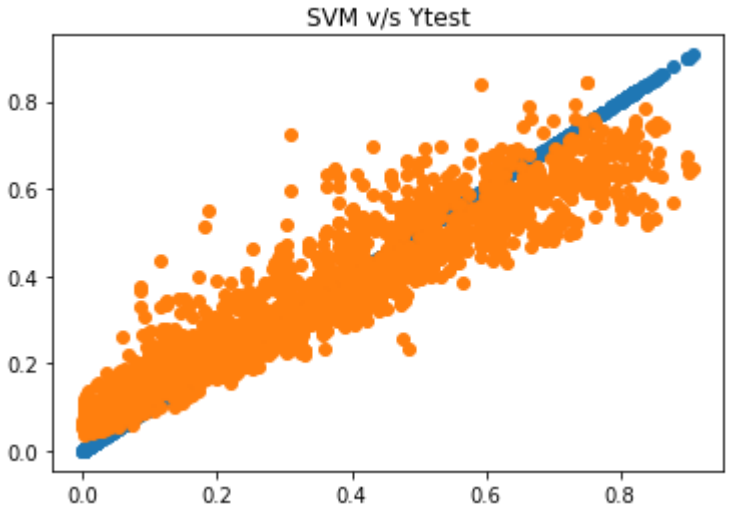


Run 2

ANN RMSE: 0.10844235468597958

SVM RMSE: 0.08664705044469487

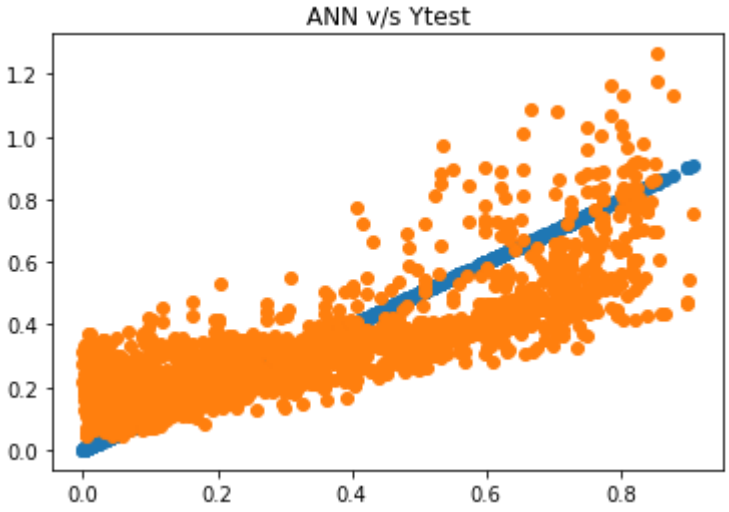


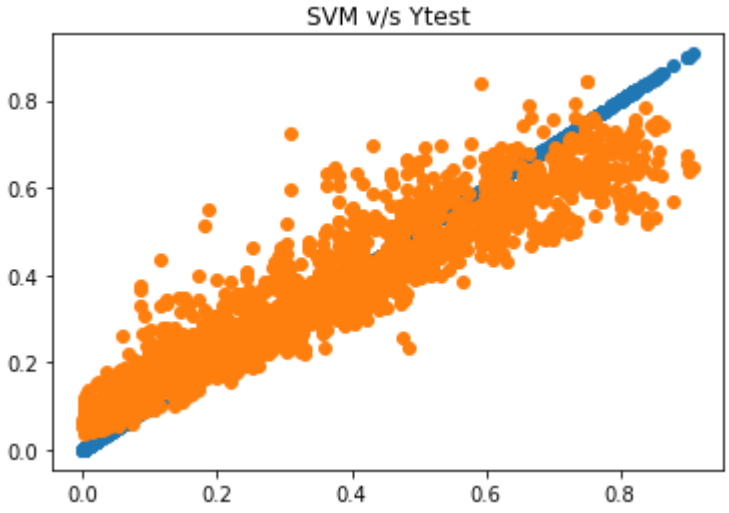


Run 3

ANN RMSE: 0.14692455241840024

SVM RMSE: 0.08664705044469487





Run 4

ANN RMSE: 0.10387968765066642

SVM RMSE: 0.08664705044469487