ANN Surrogate Modelling for Binary VLE

A Screening Task Report

Submitted by

Achyuth Sreenath Haresamudram

VIT Bhopal University

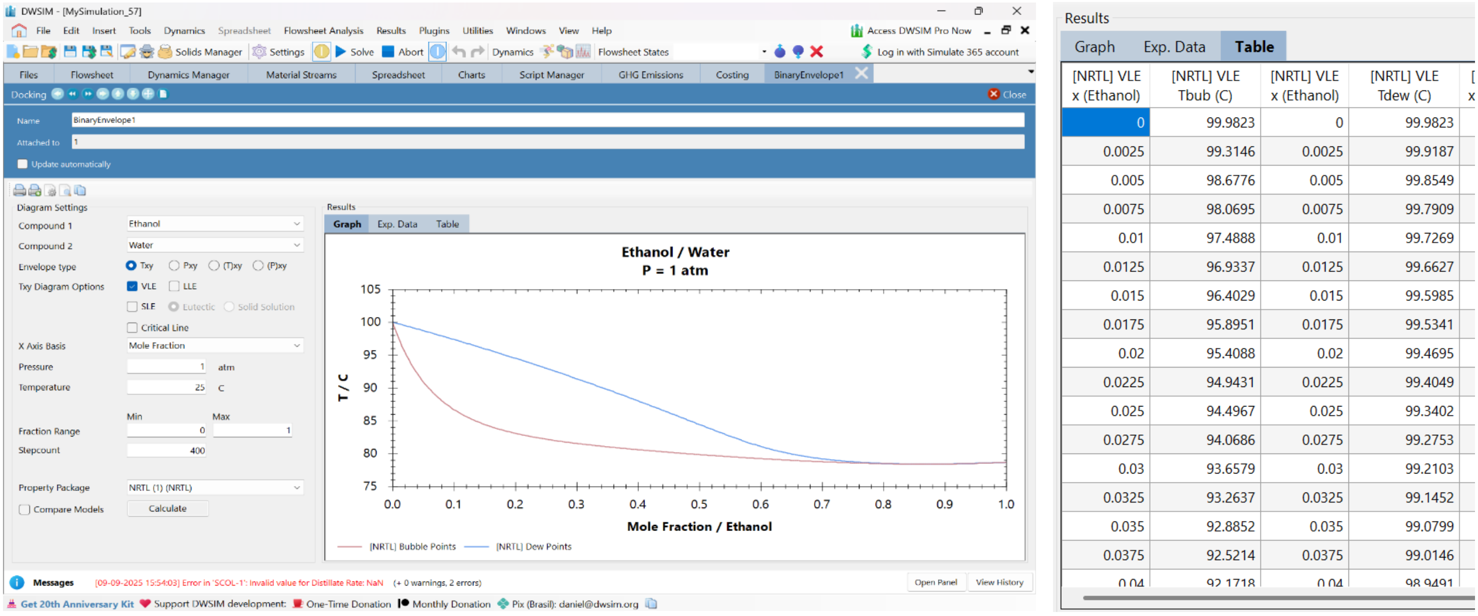
achyuth.23bai10584@vitbhopal.ac.in

Task Description:

- To build an Artificial Neural Network (ANN) model in Python to predict vapor composition in a binary azeotropic system.
- The ANN should be trained on experimental or simulated VLE data and tested on its ability to capture azeotropic behaviour.

Dataset Extraction:

- The VLE_Data.csv that is attached in the Zip file, was extracted via simulations in the **DWSIM** simulator.
- The dataset was extracted by using a flash separator and by creating a binary envelope.
  1. Selected the compounds as '**Ethanol**' and '**Water**' in the DWISM configuration wizard.
  2. Selected the '**NRTL**' property package & set the phase equilibria settings to VLE (faster) and configured the system of units to C5.
  3. Added the **flash separator** into the flowsheet and configured the material settings.
  4. Added a utility: Material Streams, Binary Phase Envelope, and 1st flowsheet object.
  5. Calculated the VLE values for Ethanol-Water Combination, in the three envelope types: **Txy, Pxy, (T)xy** and merged the results and **saved it into a .csv** file.
  6. The Screenshot of the same is attached below:



Name: BinaryEnvelope1

Attached to: 1

□ Update automatically

Diagram Settings

Compound 1: Ethanol

Compound 2: Water

Envelope type: ● Txy ○ Pxy ○ (T)xy ○ (P)xy

Txy Diagram Options: ☑ VLE □ LLE

□ SLE ● Eutectic ○ Solid Solution

□ Critical Line

X Axis Basis: Mole Fraction

Pressure: 1 atm

Temperature: 25 C

Fraction Range: Min 0 Max 1

Stepcount: 400

Property Package: NRTL (1) (NRTL)

□ Compare Models

Calculate

Messages [09-09-2025 15:54:03] Error in 'SCOL-1': Invalid value for Distillate Rate: NaN (+ 0 warnings, 2 errors)

[NRTL] VLE x (Ethanol)	[NRTL] VLE Tbub (C)	[NRTL] VLE x (Ethanol)	[NRTL] VLE Tdew (C)
0	99.9823	0	99.9823
0.0025	99.3146	0.0025	99.9187
0.005	98.6776	0.005	99.8549
0.0075	98.0695	0.0075	99.7909
0.01	97.4888	0.01	99.7269
0.0125	96.9337	0.0125	99.6627
0.015	96.4029	0.015	99.5985
0.0175	95.8951	0.0175	99.5341
0.02	95.4088	0.02	99.4695
0.0225	94.9431	0.0225	99.4049
0.025	94.4967	0.025	99.3402
0.0275	94.0686	0.0275	99.2753
0.03	93.6579	0.03	99.2103
0.0325	93.2637	0.0325	99.1452
0.035	92.8852	0.035	99.0799
0.0375	92.5214	0.0375	99.0146
0.04	92.1718	0.04	98.9491

Fig. 1: Extracting the dataset using DWSIM

Dataset Description:

- **X (mole fraction of ethanol in liquid phase):** Ranges from 0 to 1, representing a full range of compositions from pure water to pure ethanol.
- **T (Temperature):** Decreases as the ethanol concentration increases, which is expected for this azeotropic mixture.
- **P (Pressure):** The pressure data is consistently at approximately 1 atmosphere (1 atm), confirming that the data is for an isobaric system.
- **Y (mole fraction of ethanol in vapor phase):** Increases with the liquid phase concentration (X), as expected.

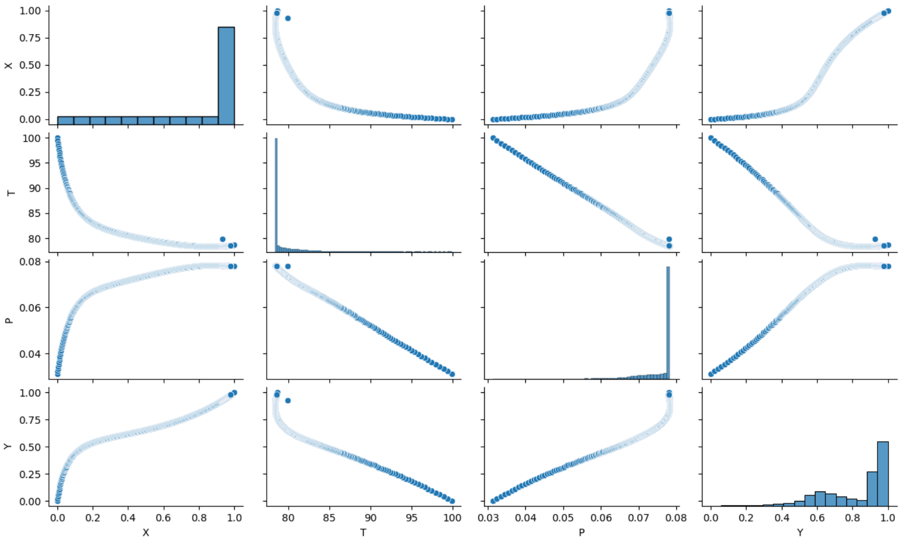


Fig. 2: VLE dataset Distribution

Methodology Followed:

The Project Design is structured around first, building an **Artificial Neural Network** (simple feedforward network) for predicting the vapour composition (**y**) of in the Ethanol-Water system based on the input of liquid phase mole composition (**X**), Temperature (**T**), and Pressure (**P).** Upon evaluation of the baseline ANN, it captured the VLE behaviour with some loss.

Then, the loss was significantly reduced by adjusting the loss function of the ANN, i.e. by building a **Physically Informed Neural Network (PINN)**.
The new loss function incorporated the modified Raoult's law that predicted the *activity coefficients*, with the same inputs, rather than the vapour pressure. Then the vapour pressure was calculated mathematically (*using formulas given later in the report*) and finally when the model was evaluated, the loss had been efficiently downsized.

Later on, the two models were compared based on **parity plots** (between the $Y_{experimental}$ v/s $Y_{predicted}$).
A **comparative study** was performed between the baseline ANN (first built) and then the Physically informed Neural Network (PINN) based on the predicted azeotropic composition, while keeping the original Raoult's law as a baseline.

Workflow:

1. Data Collection (Using DWSIM) and loading into a pandas dataframe.
2. Data preprocessing and handling the missing and duplicate values.
3. Separating the features & target, and splitting the dataset into training, testing and validation sets.
4. Standardising the data so that is, bringing features on the same scale, so the mean is around 0 and standard deviation is 1.

5. Building and Evaluating the baseline ANN (simple feedforward network using basic activation function and less epochs).
6. Building a Physics Informed Neural Network using the modified Raoult's law and predicting the activity coefficients (gamma) for the two compounds, viz. Ethanol and Water, via a custom loss.
7. Evaluation and Comparative Study by detecting the Azeotrope Composition.

Baseline ANN Design:

A simple feedforward neural network was built using 1 input layer, 1 hidden layer and an output layer consisting of basic activation functions. The Summary of the baseline ANN model is given in *Table 1.*

Baseline Neural Network Model Architecture				
Layer name	**Type**	**No. of Neurons**	**Activation Function**	**Param #**
Input layer	Flatten	3	None	0
Hidden layer 1	Dense	6	Relu	24
Output layer	Dense	1	Sigmoid	7

Table 1: Baseline model Architecture

The baseline ANN was compiled with '**Adam**' optimizer, and the loss function used was '**Mean Squared Error (MSE)**', and this was run for 50 epochs.

- Mean Squared Error (**MSE**) observed: **0.0021**
- Root Mean Square Error (**RMSE**) observed: 0.04915
- Azeotropic Composition Predicted: x = 0.92000, y_pred = 0.91380, T = 78.42
- The below figures depict the loss incurred by the Baseline ANN:

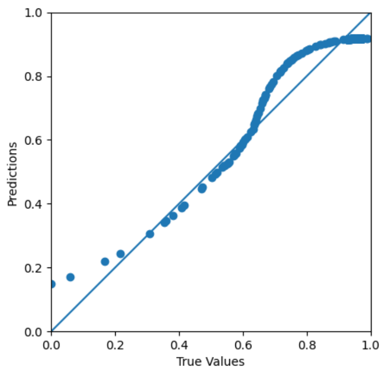


Fig. 3: Baseline model Predictions v/s True labels

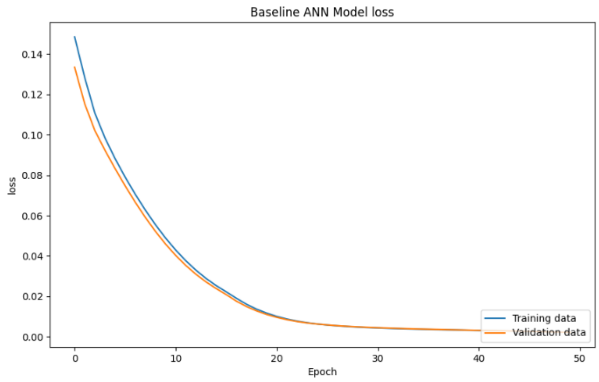


Fig. 4: Baseline model loss curve for training and validation data

Physically Informed Neural Network (PINN) Design:

Physics-Informed Neural Networks (PINNs) are a class of neural networks that incorporate physical laws and constraints into the learning process. The baseline ANN model, while simpler, may not capture the underlying physics (thermodynamic laws) of the problem as effectively as the PINN model. PINNs leverage known physical principles, such as conservation laws, modified Raoult's law (in our task), and boundary conditions, to guide the training of the model.

This integration of physics helps improve the model's accuracy, generalization, and interpretability, especially in scenarios where data may be scarce or noisy.
So, to decrease the loss, and to predict the azeotrope accurately, the need of custom loss function was necessary to build a PINN.

Building the custom loss function:

- Modified Raoult's Law:

  $$p_i = x_i \, \gamma_i \, p_i^{sat}(T)$$

  Where:

  1. p_i : partial vapor pressure of component i.
  2. x_i : liquid mole fraction.
  3. γ_i / gamma_i : activity coefficient (accounts for non-ideality).
  4. $p_i^{sat}(T)$ **:** saturation vapor pressure at temperature T, calculated using the Antoine law.
- Then we calculate the partial vapour pressure of Ethanol and water as p_1 and p_2, respectively.
- So, Total Pressure:

  $$\mathbf{P} = p_1 + p_2 = x_1 \, \gamma_1 \, p_1^{sat}(T) + (1 - x_1) \, \gamma_2 \, p_2^{sat}(T)$$

- Therefore, the vapour pressure in liquid phase:

  $$y_1 = \frac{\mathbf{p1}}{\mathbf{P}} = \frac{\mathbf{x1 \, \gamma 1 \, p1sat(T)}}{\mathbf{x1 \, \gamma 1 \, p1sat(T) + (1 - x1) \gamma 2 \, p2sat(T)}}$$

- And, the $p_i^{sat}(T)$ is calculated using the by Antoine's law: $\mathbf{Log_{10}(p_i^{sat}(T)) = A - \frac{B}{C+T}}$ (constants)
- Therefore, the PINN model will predict the activity coefficients, gamma_i (γ_i) values (instead of the vapour pressure, y_i), that is then used to calculate the y_pred or y_i (vapour pressure in liquid phase) and total pressure P.

So, the custom loss function will be:

- L(data) = $\frac{1}{N}\sum_1^N \left(y_{pred} - y_{data}\right)^2$
- L(physical) = $\frac{1}{N}\sum_1^N \left(\frac{P_{pred} - P_{data}}{P_{data}}\right)^2$
- **Combined:**

  L(total) = L(data) + L(physical)

$$\boxed{\mathbf{L(total)} = \frac{1}{N}\sum_1^N \left(y_{pred} - y_{data}\right)^2 + \lambda \, \frac{1}{N}\sum_1^N \left(\frac{P_{pred} - P_{data}}{P_{data}}\right)^2}$$

Therefore, the PINN model is built upon the above custom loss function.

The Physics informed Neural Network architecture is presented in Table 2:

Physics Informed Neural Network Model Architecture				
Layer name	**Type**	**No. of Neurons**	**Activation Function**	**Param #**
Input layer	Input	3	None	0
Hidden layer 1	Dense	64	Relu	256
Hidden layer 2	Dense	64	Relu	4160
Hidden layer 3	Dense	32	Relu	2080
Log(gamma)	Dense	2	Linear	66
Output layer	Activation	2	Exponential	0

Table 2: PINN model Architecture

The PINN was compiled with '**Adam**' optimizer, with a learning rate of 0.001 and the loss function used was '**PINN Loss** (customized as above)', and this was run for 300 epochs.
The λ (hyperparameter, weight of the physics loss) was set to 0.5.

- Mean Squared Error (**MSE**) observed: **0.00001745** or (1.745 x 10^{-5})
- Root Mean Square Error (**RMSE**) observed: 0.00310
- Azeotropic Composition Predicted: x = 0.84400, y_pred = 0.84408, T = 78.39

- Thus, the MSE has been significantly downsized and the observations are in accordance with the actual value of the azeotrope for a ethanol-water binary system.
- The below figures depict the loss incurred by the PINN:

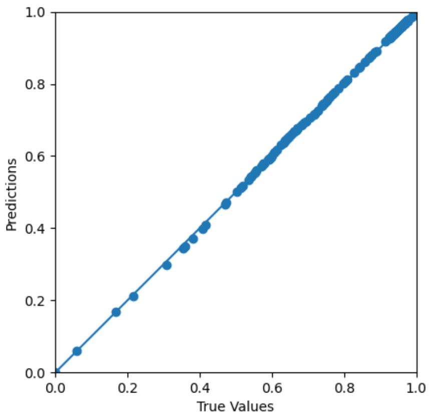


Fig. 5: PINN model Predictions v/s True labels

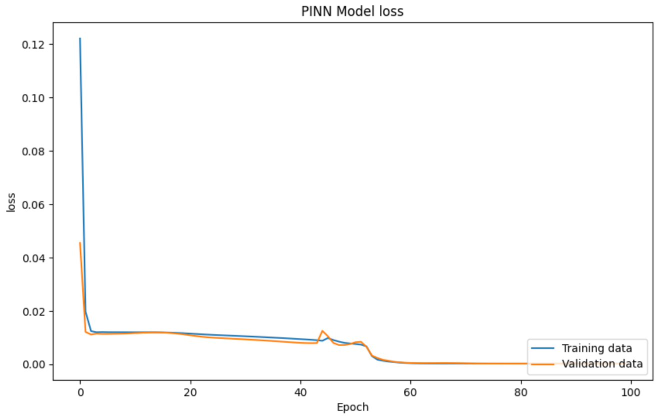


Fig. 6: PINN model loss curve for training and validation data

Results:

Both the models (ANN and PINN) were trained on the same dataset, but with different loss functions. The baseline ANN used the regular MSE function, whereas the PINN used the custom, thermodynamically altered loss function that successfully demonstrated its ability to improve prediction accuracy (reduced loss) and ensured physical consistency in modelling the VLE for a non-ideal, azeotropic system.

The summary of the Loss metric Results obtained are cited in Table 3. Additionally, the detected Azeotrope points are listed in Table 4 and the loss comparison is graphically depicted in Figure 7.

Model\Metric	MSE	RMSE
Baseline ANN	0.0021	0.04915
PINN	0.00001745	0.0031

Table 3: Loss metric result comparison

Model\Azeotrope	X – liquid mole fraction	Y – vapour mole fraction
Baseline ANN	0.920	0.9138
PINN	0.844	0.844

Table 4: Azeotrope result comparison

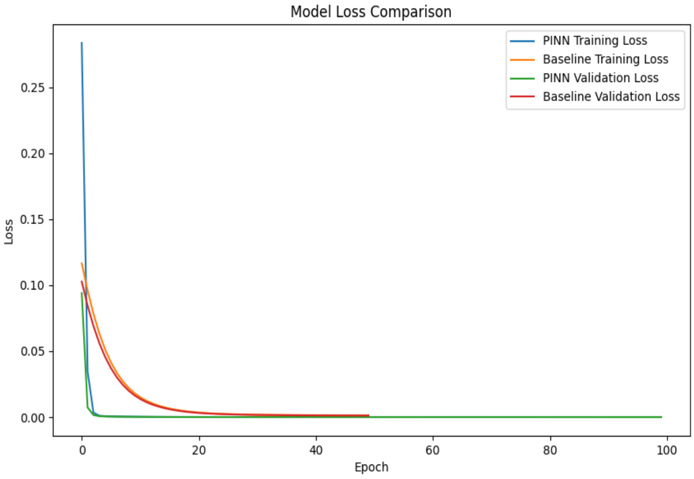


Fig. 7: ANN v/s PINN loss Comparison

Clearly,

- The **PINN model outperforms the baseline ANN model** in terms of MSE & RMSE and provides more physically consistent predictions, making it a more suitable choice for modelling VLE data.
- The PINN model also accurately predicts the azeotropic point, demonstrating its effectiveness in capturing the complex interactions in the ethanol-water binary system.
- The RMSE values for both models provide a quantitative measure of their prediction accuracy, with lower values indicating better performance.