



Modeling Membrane Distillation Experiments Using Supervised Machine Learning Methods

Presenters: *Amirhossein Alivandi*

Supervisor: *Dr. Sadeghian*

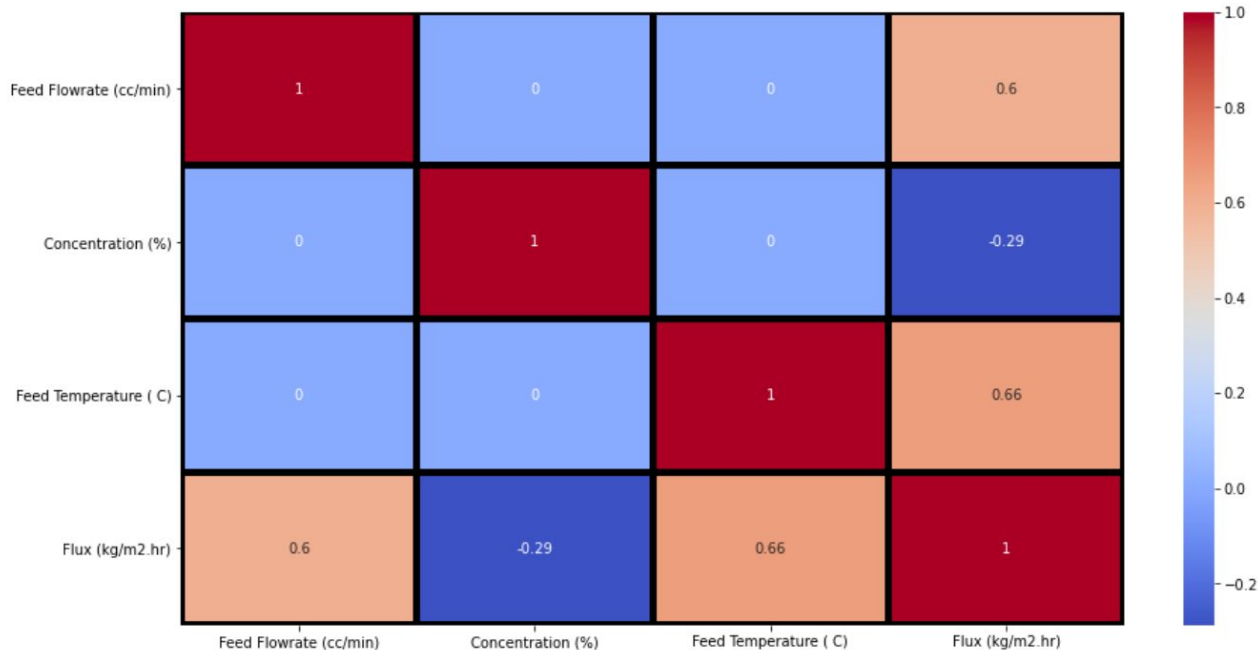
June 2022



Dataset:

- Our Variables
- Each one's Range
- Correlation Coefficient Matrix

	A	B	C	D
1	Feed Flowrate (cc/min)	Concentration (%)	Feed Temperature (C)	Flux (kg/m2.hr)
2	5	2.5	60	4.553350529
3	5	2.5	70	4.842869927
4	5	2.5	80	5.895667737
5	5	2.5	95	8.580302153
6	9	2.5	60	4.869189872
7	9	2.5	70	5.421908722
8	9	2.5	80	7.290624836
9	9	2.5	95	8.975101332
10	15	2.5	60	5.579828394
11	15	2.5	70	6.316786861
12	15	2.5	80	8.159183029
13	15	2.5	95	8.422382481
14	20	2.5	60	8.106543138
15	20	2.5	70	8.396062536
16	20	2.5	80	8.578888888



	Feed Flowrate (cc/min)	Concentration (%)	Feed Temperature (C)	Flux (kg/m2.hr)
count	40.000000	40.000000	40.000000	40.000000
mean	14.800000	3.750000	76.250000	6.880034
std	7.314194	1.265924	13.094822	1.733296
min	5.000000	2.500000	60.000000	3.921672
25%	9.000000	2.500000	67.500000	5.573248
50%	15.000000	3.750000	75.000000	6.895826
75%	20.000000	5.000000	83.750000	8.218403
max	25.000000	5.000000	95.000000	10.922777



Dataset:

- Correlation (Pearson) Matrix:
- How to calculate it ?
 - By Definition
 - Using Statistics for a large Dataset
- It's Properties:

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$$

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

- $-1 \leq \rho \leq 1$.
- If X, Y are independent, then $\rho = 0$.
- $|\rho| = 1$ if and only if $X - \mathbb{E}[X] = c(Y - \mathbb{E}[Y])$.
- $\rho(aX + b, Y) = \text{sign}(a)\rho(X, Y)$.



History:

- ANN Modeling :
- MVLRA :

جدول (۴-۴) نتایج مدل رگرسیون چند متغیر MVLRA

Regression Statistics

Multiple R 0.928756

R Square 0.862588

Adjusted R Square 0.859042

Standard Error 0.646575

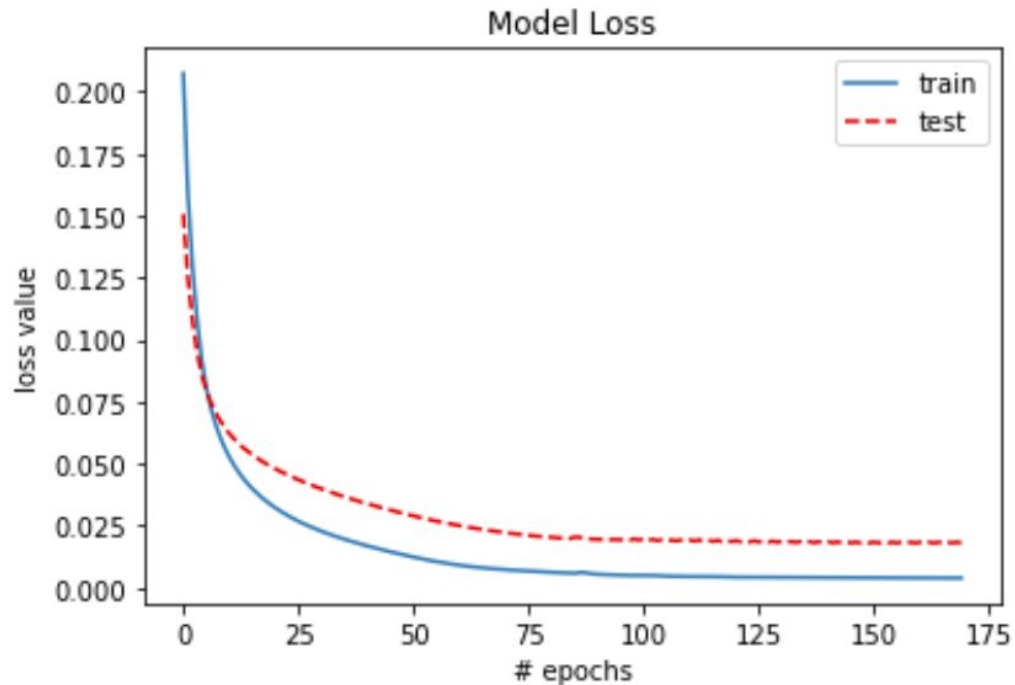
جدول (۳-۴) ساختار بررسی شده در شبکه عصبی

ساختار مدل	الگوریتم آموزش	RMSE	R ²
5-4-1	LM	0.02	0.992



ANN Modeling:

- Best Results :



R2_test	R2_train	Optimizer	Structure	Activation Function
0.7601	0.9074	Adam	3-13-13-1	ReLU
0.7758	0.8732	Adam	3-5-4-1	Sigmoid
0.7363	0.9017	Adam	3-6-5-1	Softmax
0.8558	0.8582	Adam	3-8-1	Softplus
0.8637	0.8538	Adam	3-7-1	Softsign
0.8027	0.8928	Adam	3-4-4-1	Tanh
0.8086	0.8832	RMSprop	3-5-4-1	Softplus
0.8195	0.8822	Nadam	3-5-4-1	Softplus
0.7392	0.9390	RMSprop	3-5-4-1	Softsign
0.7753	0.9028	Nadam	3-5-4-1	Softsign
0.9382	0.8462	RMSprop	3-5-4-1	Tanh



Regression Modeling:

- Result :
 - MAE: 0.09446
 - MSE: 0.0136
 - RMSE: 0.11661

GLSAR Regression Results

```

=====
Dep. Variable:          0      R-squared:                0.892
Model:                  GLSAR  Adj. R-squared:           0.878
Method:                 Least Squares  F-statistic:            63.30
Date:                   Fri, 24 Jun 2022  Prob (F-statistic):     2.88e-11
Time:                   06:34:38      Log-Likelihood:         30.794
No. Observations:      27      AIC:                   -53.59
Df Residuals:          23      BIC:                   -48.40
Df Model:               3
Covariance Type:       nonrobust
=====

```

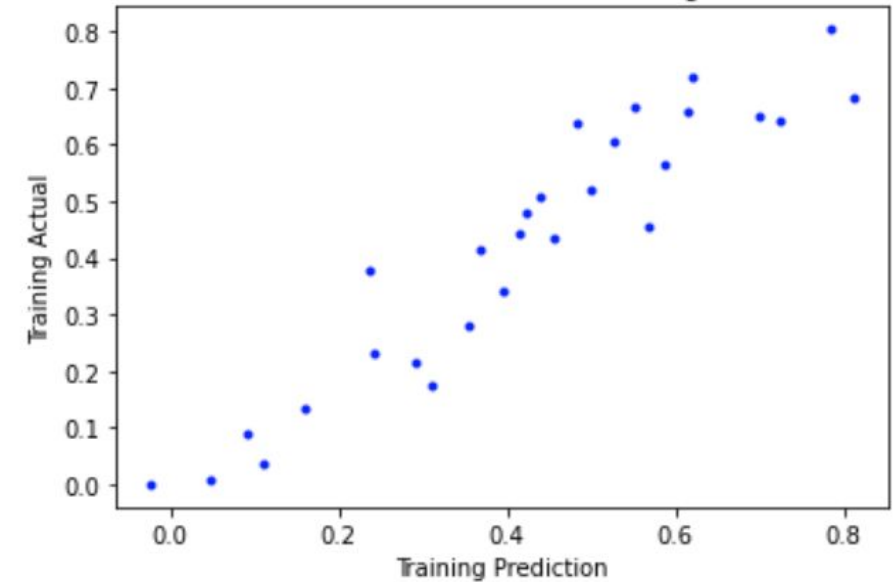
	coef	std err	t	P> t	[0.025	0.975]
const	0.0933	0.036	2.562	0.017	0.018	0.169
0	0.3602	0.047	7.661	0.000	0.263	0.458
1	-0.1215	0.033	-3.643	0.001	-0.191	-0.053
2	0.4457	0.049	9.103	0.000	0.344	0.547

```

=====
Omnibus:                0.837  Durbin-Watson:          2.434
Prob(Omnibus):          0.658  Jarque-Bera (JB):       0.807
Skew:                   0.198  Prob(JB):               0.668
Kurtosis:               2.251  Cond. No.               4.67
=====

```

Actual Vs. Predicted Training



OLS Regression Results

```

=====
Dep. Variable:          0      R-squared:                0.888
Model:                  OLS    Adj. R-squared:           0.874
Method:                 Least Squares  F-statistic:            63.69
Date:                   Sat, 28 May 2022  Prob (F-statistic):     1.42e-11
Time:                   15:22:24      Log-Likelihood:         31.918
No. Observations:      28      AIC:                   -55.84
Df Residuals:          24      BIC:                   -50.51
Df Model:               3
Covariance Type:       nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	0.0906	0.036	2.502	0.020	0.016	0.165
0	0.3462	0.044	7.786	0.000	0.254	0.438
1	-0.1130	0.032	-3.530	0.002	-0.179	-0.047
2	0.4613	0.046	10.039	0.000	0.366	0.556

```

=====
Omnibus:                0.661  Durbin-Watson:          2.345
Prob(Omnibus):          0.719  Jarque-Bera (JB):       0.710
Skew:                   0.169  Prob(JB):               0.701
Kurtosis:               2.297  Cond. No.               4.58
=====

```

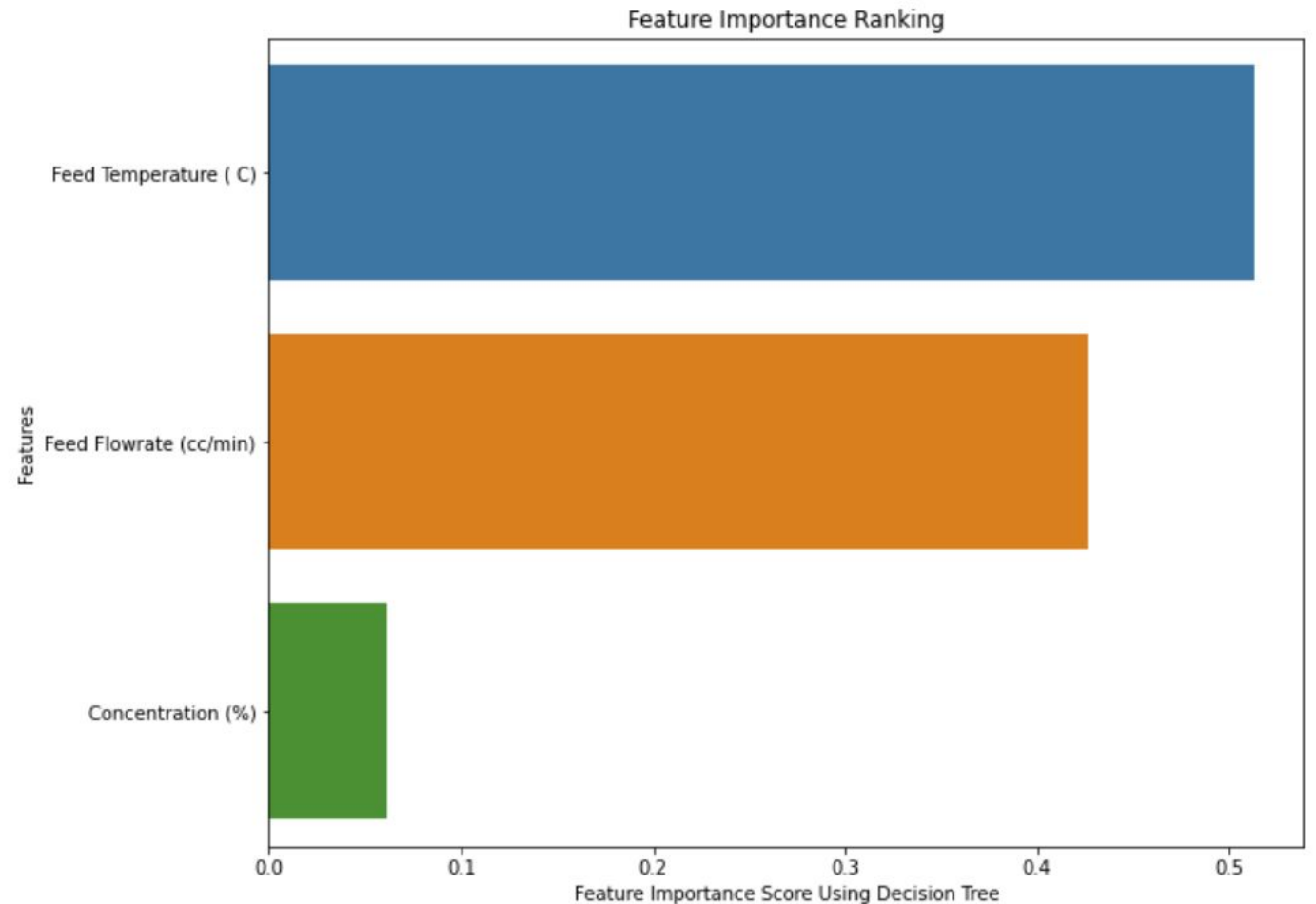


DT Modeling:

- Result :

Training Data $R^2 = 0.9247$ $R = 0.9616$
Testing Data $R^2 = 0.5276$ $R = 0.7264$

MAE: 1.1069
MSE: 1.79879
RMSE: 1.34119





RF & ET Modeling:

- RF's Best Results :

Training Data $R^2 = 0.9211$ $R = 0.9597$
Testing Data $R^2 = 0.8397$ $R = 0.9164$

MAE: 0.59831
MSE: 0.47951
RMSE: 0.69247

Training Data $R^2 = 0.9212$ $R = 0.9598$
Testing Data $R^2 = 0.8009$ $R = 0.8949$

MAE: 0.69187
MSE: 0.61326
RMSE: 0.78311

- ET's Best Results :

Training Data $R^2 = 0.9594$ $R = 0.9795$
Testing Data $R^2 = 0.889$ $R = 0.9429$

MAE: 0.4599
MSE: 0.35636
RMSE: 0.59696



GB, XGB & AdaBoost Modeling:

- GB's Result :

Training Data $R^2 = 0.9816$ $R = 0.9907$
Testing Data $R^2 = 0.8615$ $R = 0.9282$

MAE: 0.46682
MSE: 0.32359
RMSE: 0.56885

- XGB's Result :

Training Data $R^2 = 0.9899$ $R = 0.9949$
Testing Data $R^2 = 0.8889$ $R = 0.9428$

MAE: 0.4081
MSE: 0.26835
RMSE: 0.51802

- AdaBoost's Result :

Training Data $R^2 = 0.9815$ $R = 0.9907$
Testing Data $R^2 = 0.8766$ $R = 0.9363$

MAE: 0.60792
MSE: 0.64919
RMSE: 0.80572



Thanks for your
time and attention