VIETNAM NATIONAL UNIVERSITY HO CHI MINH CITY
HO CHI MINH CITY UNIVERSITY OF TECHNOLOGY

RESEARCH TOPIC: DATA ANALYTICS

MINI PROJECT

Linear regression and Polynomial regression model for chemical experiments

Lecturer: Assoc. Prof. Tran Minh Quang

Group 7:

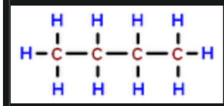
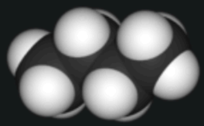
Đinh Thanh Phong	2270243
Trần Ngọc Phụng	2270123
Thái Học Phú	2270183

Table of Contents

1. Project introduction and objectives:

Project 1: find the relationship between “time” and “humidity” by linear regression model. aims to optimize drying for materials, thereby applying to practical industries such as drying paper and other materials.

Project 2: find the relationship between "temperature" and "pressure" of Butanol gas by using a polynomial regression model. Thereby optimizing productivity and safety in the production of Butanol gas.

Butan	
	
Tổng quan	
Công thức hóa học	C_4H_{10}
SMILES	CCCC
Phân tử gam	58,08 g/mol
Bề ngoài	chất khí không màu
số CAS	[106-97-8]
Thuộc tính	
Tỷ trọng và pha	12.52 g/l, khí
Độ hoà tan trong nước	6,1 mg/100 ml ở 20 °C)
Nhiệt độ nóng chảy	- 138,3 °C (134,9 K)
Nhiệt độ sôi	- 0,5 °C (272,7 K)

Nguy hiểm	
MSDS	MSDS ngoài
Phân loại của EU	Rất dễ cháy (**F+**)
NFPA 704	NFPA 704 ***"Biểu đồ cháy"*** 
Nguy hiểm	R12
An toàn	S2, S9, S16
Điểm bốc cháy	- 60 °C
Nhiệt độ tự bốc cháy	287 °C
Giới hạn nổ	1,8–8,4%
Số RTECS	

2. Linear regression for convection drying experiment:

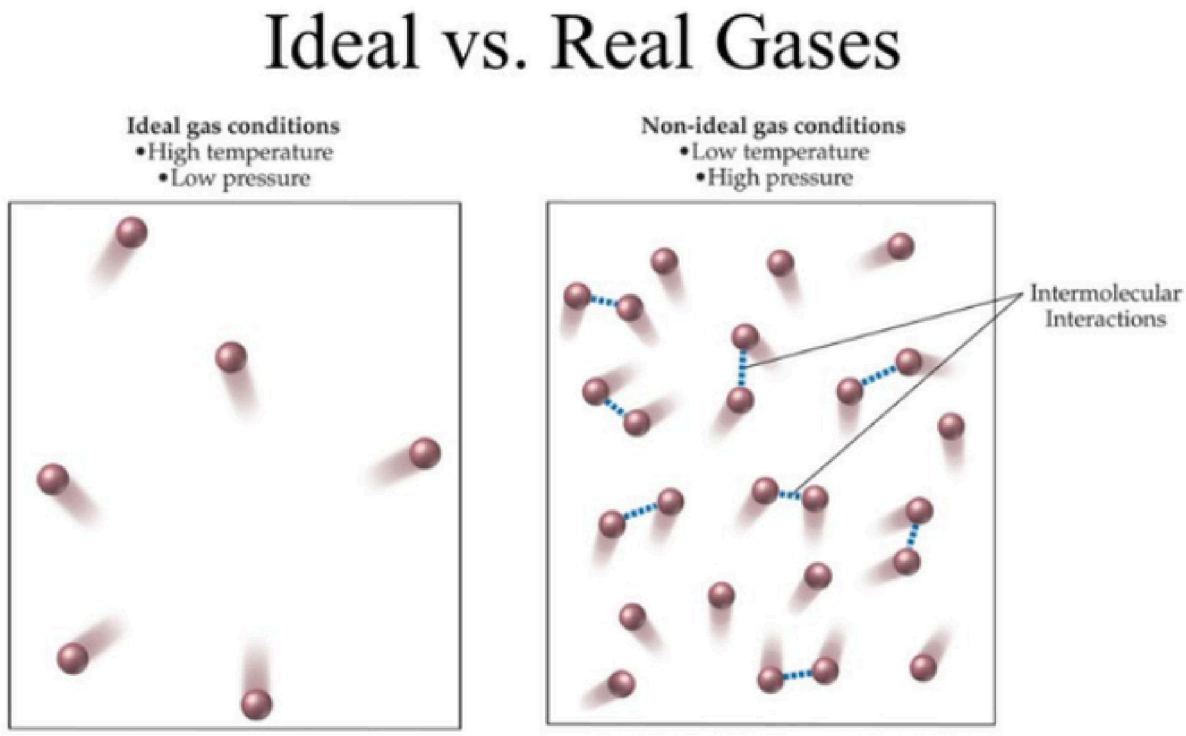
3. Polynomial regression for correlation between vapor pressure and temperature of Butanol:

3.1) Difference Between Ideal Gas and Real Gas

Real gas and Ideal gas. As the particle size of an ideal gas is extremely small and the mass is almost zero and no volume Ideal gas is also considered a point mass. The molecules of real gas occupy space though they are small particles and also have volume.

Ideal gas: An ideal gas is defined as a gas that obeys gas laws at all conditions of pressure and temperature. Ideal gasses have velocity and mass. They do not have volume. When compared to the total volume of the gas the volume occupied by the gas is negligible. It does not condense and does not have triple points.

Real gas: A real gas is defined as a gas that does not obey gas laws at all standard pressure and temperature conditions. When the gas becomes massive and voluminous it deviates from its ideal behavior. Real gasses have velocity, volume and mass. When they are cooled to their boiling point, they liquefy. When compared to the total volume of the gas the volume occupied by the gas is not negligible.



To make you understand how ideal gas and real gas are different from each other, here are some of the major differences between ideal gas and real gas:

Ideal gas	Real gas
Ideal gas obeys all gas laws under all conditions of pressure and temperature.	Real gas obeys gas laws only at conditions of low pressure and high temperature. They obey Vanderwaal's real gas equation
The molecules collide with each other elastically.	The molecules collide with each other inelastically.
The volume occupied by the molecules is negligible as compared to the total volume.	The volume occupied by molecules is not negligible as compared to total volume.
There are no intermolecular forces of attraction.	Either attractive or repulsive forces are present between the particles.
It is a hypothetical gas.	It exists in nature around us.
It has high pressure	It has a pressure correction term in its equation and the actual pressure is less than ideal gas.
Obeys PV = nRT	Obeys $(P + \frac{an^2}{V^2})(V - nb) = nRT$

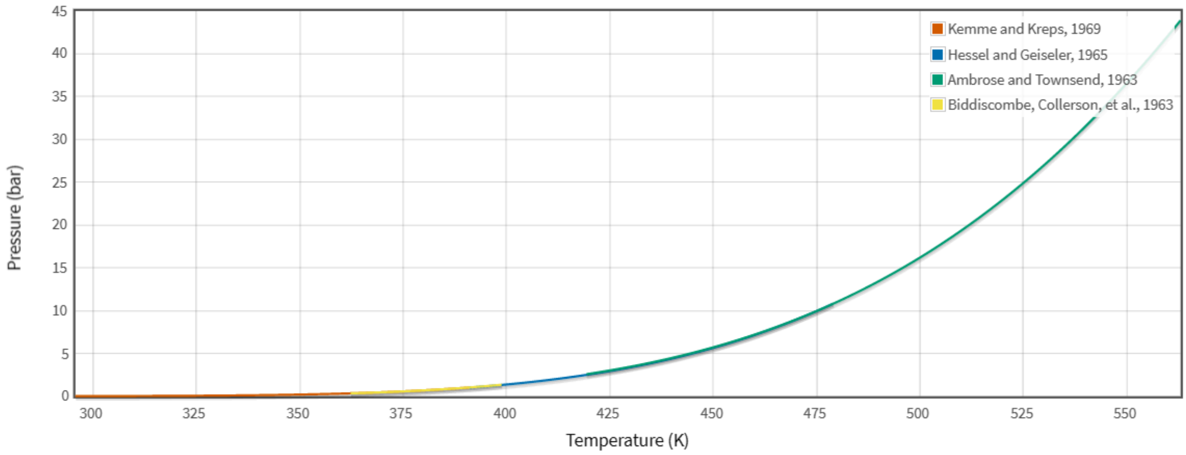
3.2) Real gas: 1-Butanol

- Formula: $C_4H_{10}O$
- Other names: Butyl alcohol; n-Butan-1-ol; n-Butanol; n-Butyl alcohol; Butyl hydroxide; CCS 203; Hemostyp; Methylolpropane; Propylcarbinol; n-C4H9OH; Butanol; Butan-1-ol; 1-Hydroxybutane; Alcool butylique; Butanolo; Butylowy alkohol; Butyric alcohol; Propylmethanol; Butanolen; 1-Butyl alcohol; Rcra waste number U031; Butanol-1; NSC 62782
- Antoine Equation Parameters:

$$\log10(P) = A - (B / (T + C))$$

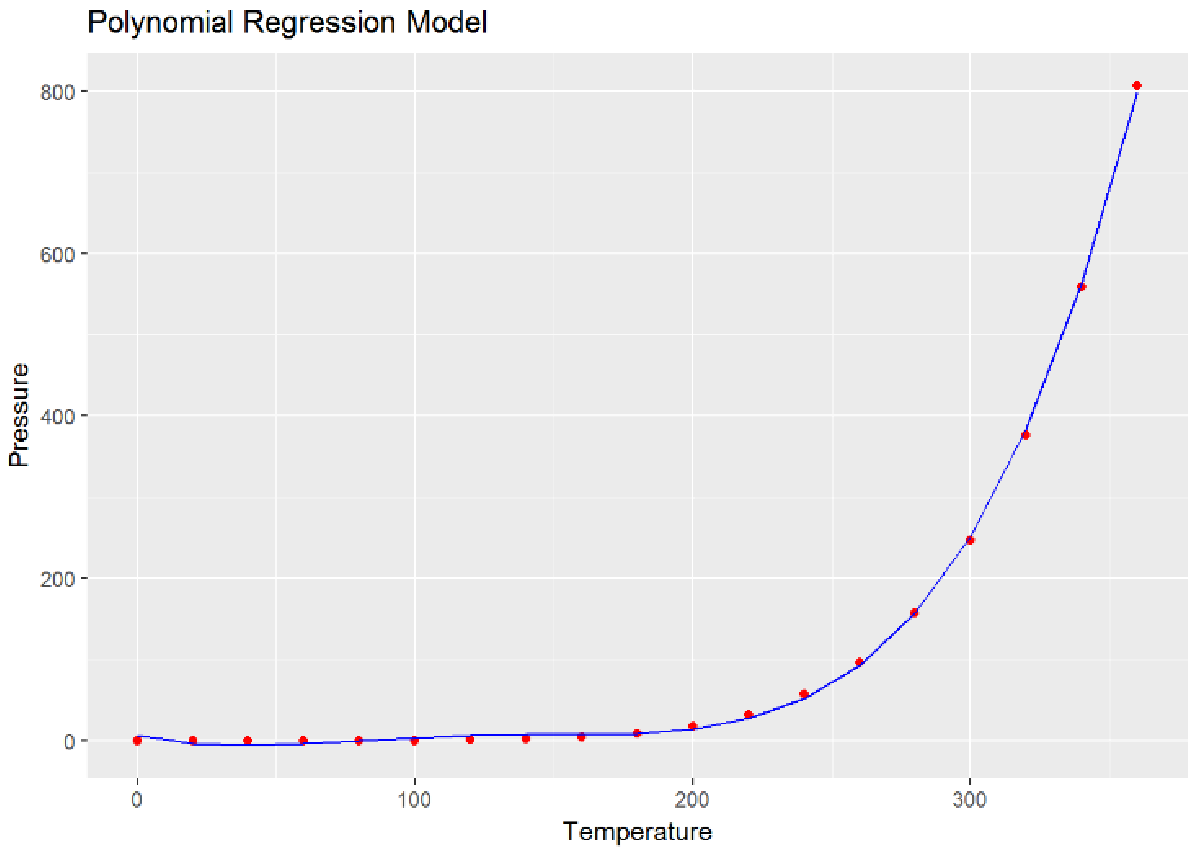
P = vapor pressure (bar)
T = temperature (K)

Temperature (K)	A	B	C	Reference	Comment
295.8 - 391.0	4.54607	1351.555	-93.34	Kemme and Kreps, 1969	
391. - 479.	4.39031	1254.502	-105.246	Hessel and Geiseler, 1965	Coefficents calculated by NIST from author's data.
419.34 - 562.98	4.42921	1305.001	-94.676	Ambrose and Townsend, 1963	Coefficents calculated by NIST from author's data.
362.36 - 398.84	4.50393	1313.878	-98.789	Biddiscombe, Collerson, et al., 1963	Coefficents calculated by NIST from author's data.



3.3) Introduction Polynomial Regression using R

By using R, the report on https://rpubs.com/anup_jana/polynomial shows us the polynomial model that was built by 19 observations. of 2 variables.



```
summary(poly_reg1) # check the summary of polynomial model
```

```
##
## Call:
## lm(formula = pressure ~ ., data = poly_pressure1)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.1989 -4.2112  0.2224  4.0172  7.0729
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)   6.453e+00  4.645e+00   1.389 0.186418
## temperature  -7.992e-01  1.893e-01  -4.223 0.000852 ***
## temperature2  1.588e-02  2.226e-03   7.135 5.06e-06 ***
## temperature3 -1.052e-04  9.415e-06 -11.179 2.31e-08 ***
## temperature4  2.341e-07  1.297e-08  18.056 4.28e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.38 on 14 degrees of freedom
## Multiple R-squared:  0.9996, Adjusted R-squared:  0.9994
## F-statistic:  7841 on 4 and 14 DF,  p-value: < 2.2e-16
```

You can see from the summary of the model that all transformed temperature variables are significant and R2 of the model is 99.96%.

Inspired by the report, we want to build another polynomial model that uses Python language and was built by a larger dataset, 86 observations as the section below.

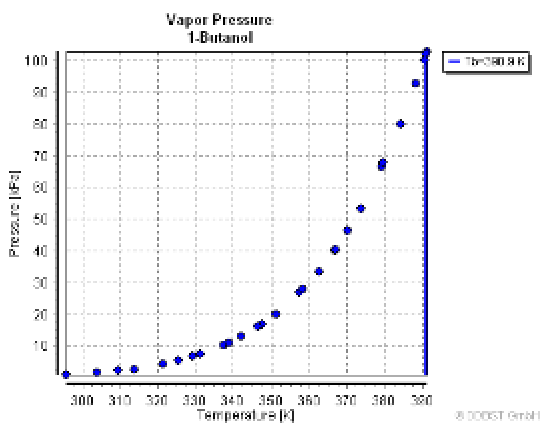
3.4) The 1-Butanol experimental data

The experimental data shown in these pages are freely available and have been published already in the DDB Explorer Edition.

Component

Formula	Molar Mass	CAS Registry Number	Name
$C_4H_{10}O$	74.123	71-36-3	1-Butanol

Diagrams



See larger image

T [K]	P [kPa]	State
295.75	0.7333	Vapor-Liquid
298.13	0.905	Vapor-Liquid
303.15	1.277	Vapor-Liquid
304.05	1.3732	Vapor-Liquid
308.18	1.809	Vapor-Liquid
309.35	1.973	Vapor-Liquid
313.85	2.613	Vapor-Liquid
321.35	4.133	Vapor-Liquid
323.25	4.593	Vapor-Liquid
325.55	5.280	Vapor-Liquid
329.44	6.540	Vapor-Liquid
331.45	7.373	Vapor-Liquid
333.27	8.091	Vapor-Liquid
337.51	10.100	Vapor-Liquid
338.95	10.852	Vapor-Liquid
342.29	12.910	Vapor-Liquid
343.45	13.812	Vapor-Liquid
346.65	16.185	Vapor-Liquid
347.69	16.850	Vapor-Liquid
351.13	19.870	Vapor-Liquid
357.54	26.700	Vapor-Liquid
358.25	27.731	Vapor-Liquid
362.36	33.045	Vapor-Liquid
362.59	33.370	Vapor-Liquid
366.81	39.974	Vapor-Liquid

366.85	40.030	Vapor-Liquid
366.95	40.463	Vapor-Liquid
370.31	46.230	Vapor-Liquid
370.51	46.601	Vapor-Liquid
373.70	53.090	Vapor-Liquid
373.89	53.450	Vapor-Liquid
374.32	54.436	Vapor-Liquid
376.79	59.932	Vapor-Liquid
379.37	66.280	Vapor-Liquid
379.52	66.632	Vapor-Liquid
379.65	67.594	Vapor-Liquid
382.04	73.331	Vapor-Liquid
383.35	77.140	Vapor-Liquid
384.31	79.856	Vapor-Liquid
384.34	79.930	Vapor-Liquid
386.58	86.807	Vapor-Liquid
388.41	92.750	Vapor-Liquid
388.47	92.995	Vapor-Liquid
390.54	100.142	Vapor-Liquid
390.57	100.210	Vapor-Liquid
390.95	102.125	Vapor-Liquid
391.30	102.830	Vapor-Liquid
392.34	106.705	Vapor-Liquid
394.09	113.412	Vapor-Liquid
395.71	119.945	Vapor-Liquid
397.31	126.628	Vapor-Liquid
398.15	123.100	Vapor-Liquid
398.84	133.322	Vapor-Liquid
419.34	254.731	Vapor-Liquid
423.15	269.200	Vapor-Liquid
429.11	335.690	Vapor-Liquid
433.77	381.995	Vapor-Liquid
439.24	439.041	Vapor-Liquid
439.28	439.447	Vapor-Liquid
443.97	492.946	Vapor-Liquid
448.15	515.300	Vapor-Liquid
448.63	554.957	Vapor-Liquid
459.75	719.306	Vapor-Liquid
462.64	764.497	Vapor-Liquid
470.31	905.744	Vapor-Liquid
472.55	947.794	Vapor-Liquid
473.15	925.700	Vapor-Liquid

480.96	1128.560	Vapor-Liquid
482.32	1158.750	Vapor-Liquid
490.87	1346.910	Vapor-Liquid
492.30	1404.470	Vapor-Liquid
498.15	1518.300	Vapor-Liquid
502.06	1683.210	Vapor-Liquid
502.47	1692.740	Vapor-Liquid
512.82	2023.360	Vapor-Liquid
513.06	2044.430	Vapor-Liquid
522.92	2404.750	Vapor-Liquid
523.15	2372.300	Vapor-Liquid
523.20	2412.550	Vapor-Liquid
532.85	2818.460	Vapor-Liquid
533.23	2827.880	Vapor-Liquid
542.77	3283.940	Vapor-Liquid
548.15	3567.100	Vapor-Liquid
550.64	3694.710	Vapor-Liquid
556.89	4053.610	Vapor-Liquid
562.98	4413.110	Vapor-Liquid

3.5) Application Polynomial Regression using Python

3.5.1) Build model:

Steps to set up the model:

- The model was built on Google Colab.
- The source code:

```
import numpy as np
import pandas as pd

import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('VaporPressureofButanol.csv')
X = df.iloc[:,0]
X = X.to_numpy()

X.shape

y = df.iloc[:,1]
```

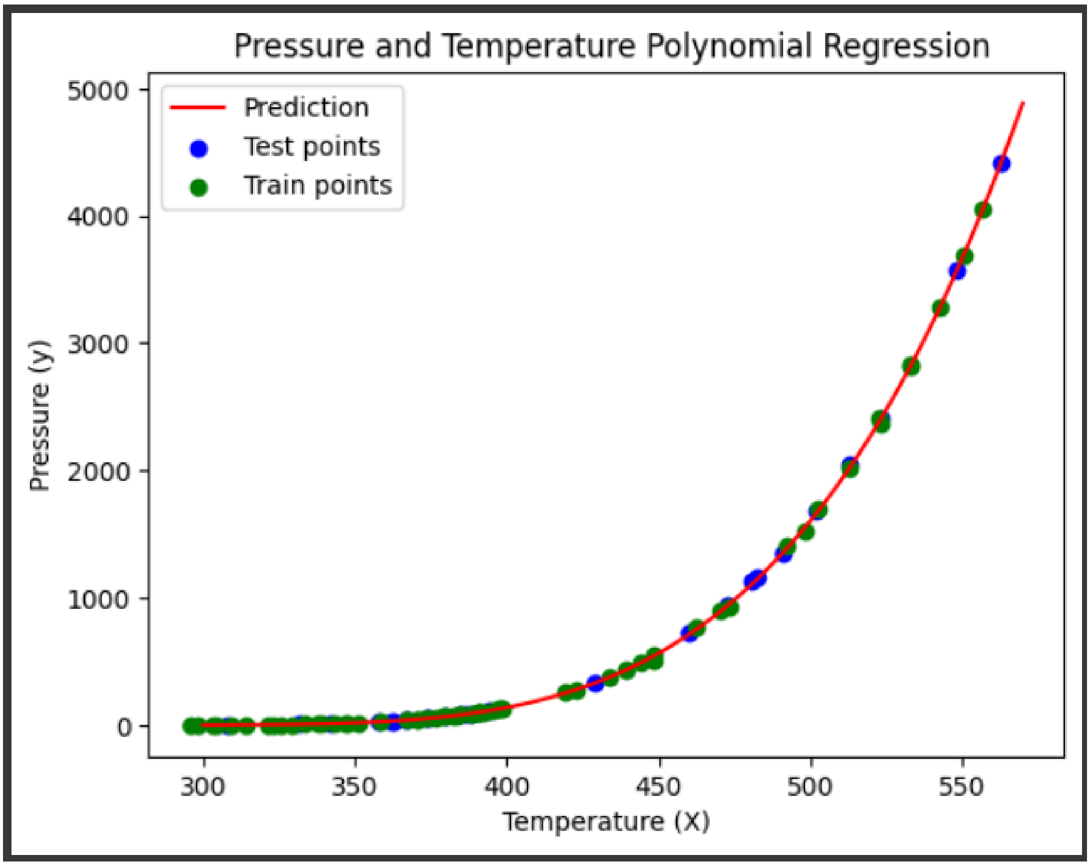
```
y = y.to_numpy()

y.shape

plt.figure()
plt.scatter(X, y, c='b')
plt.xlabel("data")
plt.ylabel("target/label")
plt.title(" All data points")
plt.show()

# split to train and test
from sklearn.model_selection import train_test_split

X = X.reshape(-1,1)
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.3)

print("X_train shape:", X_train.shape)
print("y_train shape:", y_train.shape)
print("X_test shape:", X_test.shape)
print("y_test shape:", y_test.shape)

from sklearn.linear_model import LinearRegression

model = LinearRegression()
model.fit(X_train, y_train)

N_draw = 100
Xmin=300
Xmax=570
X_draw = np.linspace(Xmin, Xmax, N_draw)
y_draw = model.predict(X_draw.reshape(-1,1))

plt.figure()
plt.scatter(X_test.ravel(), y_test, c='b', label = 'Test
points')
plt.scatter(X_train.ravel(), y_train, c='g', label = 'Train
points')

plt.plot(X_draw, y_draw, '-r', label="Prediction")
plt.xlabel("data (x)")
```

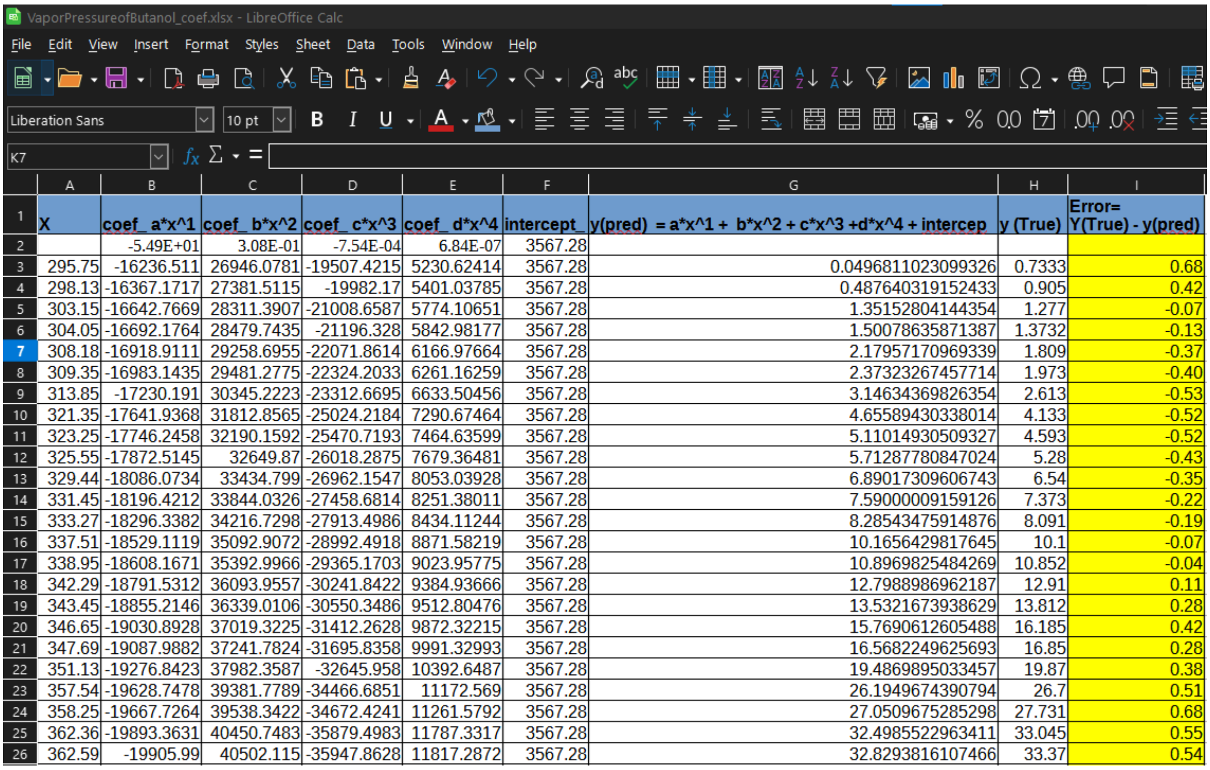
```
plt.ylabel("target/label (y)")
plt.title(" All data points")
plt.legend()
plt.show()

model.coef_

model.intercept_

import sklearn.metrics as metrics

y_pred = model.predict(X_test)
mae = metrics.mean_absolute_error(y_test, y_pred)
mse = metrics.mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)


print("MIN : MAX : MEDIAN = {:<5.2f} : {:<5.2f}: :
{:<5.2f}".format(np.abs(y_test - y_pred).min(),
                                                  np.abs(y_test -
y_pred).max(),

np.median(np.abs(y_test - y_pred)) ))
print("MSE: {:<5.2f}".format(mse))
print("RMSE: {:<5.2f}".format(rmse))
print("MAE: {:<5.2f}".format(mae))

y_test

from sklearn.preprocessing import PolynomialFeatures

def feature_extractor(X, degree=2, interaction_only=False,
include_bias=True):
  transformer = PolynomialFeatures(degree=degree,

interaction_only=interaction_only,
                                   include_bias=include_bias)
  return transformer.fit_transform(X)

X_train_trans = feature_extractor(X_train, degree=4)
X_test_trans = feature_extractor(X_test, degree=4)
```

```python
print("X_train.shape: ", X_train.shape)
print("X_test.shape: ", X_test.shape)

print("X_train_trans.shape: ", X_train_trans.shape)
print("X_test_trans.shape: ", X_test_trans.shape)

improved_model = LinearRegression()
improved_model.fit(X_train_trans, y_train)

N_draw = 100
X_draw = np.linspace(Xmin, Xmax, N_draw)

X_draw_trans = feature_extractor(X_draw.reshape(-1,1), degree=4)
y_draw = improved_model.predict(X_draw_trans)


plt.figure()
plt.plot(X_draw, y_draw, '-r', label="Prediction")

plt.scatter(X_test.ravel(), y_test, c='b', label = 'Test
points')
plt.scatter(X_train.ravel(), y_train, c='g', label = 'Train
points')

plt.xlabel("Temperature (X)")
plt.ylabel("Pressure (y)")
plt.title(" Pressure and Temperature Polynomial Regression")
plt.legend()
plt.show()

X_test_trans.shape

import sklearn.metrics as metrics

y_pred = improved_model.predict(X_test_trans) # X_test =>
X_test_trans
mae = metrics.mean_absolute_error(y_test, y_pred)
mse = metrics.mean_squared_error(y_test, y_pred)
rmse = np.sqrt(mse)
R2 = metrics.r2_score(y_test, y_pred)
```

```
print("MIN : MAX : MEDIAN = {:<5.2f} : {:<5.2f}: :
{:<5.2f}".format(np.abs(y_test - y_pred).min(),
                                                np.abs(y_test -
y_pred).max(),

np.median(np.abs(y_test - y_pred)) ))
print("MSE: {:<5.2f}".format(mse))
print("RMSE: {:<5.2f}".format(rmse))
print("MAE: {:<5.2f}".format(mae))
print("R2: {:<5.2f}".format(R2))

error = abs(y_test - y_pred)
error

plt.figure()
plt.hist(error, bins=50, density=True)
plt.xlabel("error")
plt.ylabel("Frequency")
plt.title("Distribution of errors")
plt.show()

improved_model.coef_

improved_model.intercept_
```

3.5.2) Model evaluation:





	A	B	C	D	E	F	G	H	I
1	X	coef_ a*x^1	coef_ b*x^2	coef_ c*x^3	coef_ d*x^4	intercept_	y(pred) = a*x^1 + b*x^2 + c*x^3 +d*x^4 + intercep	y (True)	Error= Y(True) - y(pred)
2		-5.49E+01	3.08E-01	-7.54E-04	6.84E-07	3567.28			
3	295.75	-16236.511	26946.0781	-19507.4215	5230.62414	3567.28	0.0496811023099326	0.7333	0.68
4	298.13	-16367.1717	27381.5115	-19982.17	5401.03785	3567.28	0.487640319152433	0.905	0.42
5	303.15	-16642.7669	28311.3907	-21008.6587	5774.10651	3567.28	1.35152804144354	1.277	-0.07
6	304.05	-16692.1764	28479.7435	-21196.328	5842.98177	3567.28	1.50078635871387	1.3732	-0.13
7	308.18	-16918.9111	29258.6955	-22071.8614	6166.97664	3567.28	2.17957170969339	1.809	-0.37
8	309.35	-16983.1435	29481.2775	-22324.2033	6261.16259	3567.28	2.37323267457714	1.973	-0.40
9	313.85	-17230.191	30345.2223	-23312.6695	6633.50456	3567.28	3.14634369826354	2.613	-0.53
10	321.35	-17641.9368	31812.8565	-25024.2184	7290.67464	3567.28	4.65589430338014	4.133	-0.52
11	323.25	-17746.2458	32190.1592	-25470.7193	7464.63599	3567.28	5.11014930509327	4.593	-0.52
12	325.55	-17872.5145	32649.87	-26018.2875	7679.36481	3567.28	5.71287780847024	5.28	-0.43
13	329.44	-18086.0734	33434.799	-26962.1547	8053.03928	3567.28	6.89017309606743	6.54	-0.35
14	331.45	-18196.4212	33844.0326	-27458.6814	8251.38011	3567.28	7.59000009159126	7.373	-0.22
15	333.27	-18296.3382	34216.7298	-27913.4986	8434.11244	3567.28	8.28543475914876	8.091	-0.19
16	337.51	-18529.1119	35092.9072	-28992.4918	8871.58219	3567.28	10.1656429817645	10.1	-0.07
17	338.95	-18608.1671	35392.9966	-29365.1703	9023.95775	3567.28	10.8969825484269	10.852	-0.04
18	342.29	-18791.5312	36093.9557	-30241.8422	9384.93666	3567.28	12.7988986962187	12.91	0.11
19	343.45	-18855.2146	36339.0106	-30550.3486	9512.80476	3567.28	13.5321673938629	13.812	0.28
20	346.65	-19030.8928	37019.3225	-31412.2628	9872.32215	3567.28	15.7690612605488	16.185	0.42
21	347.69	-19087.9882	37241.7824	-31695.8358	9991.32993	3567.28	16.5682249625693	16.85	0.28
22	351.13	-19276.8423	37982.3587	-32645.958	10392.6487	3567.28	19.4869895033457	19.87	0.38
23	357.54	-19628.7478	39381.7789	-34466.6851	11172.569	3567.28	26.1949674390794	26.7	0.51
24	358.25	-19667.7264	39538.3422	-34672.4241	11261.5792	3567.28	27.0509675285298	27.731	0.68
25	362.36	-19893.3631	40450.7483	-35879.4983	11787.3317	3567.28	32.4985522963411	33.045	0.55
26	362.59	-19905.99	40502.115	-35947.8628	11817.2872	3567.28	32.8293816107466	33.37	0.54

The result show that:

- The R2 value is 1.00
- The model plot is fitting with the data.

- The model formula is:

y(pred) = a*x^1 + b*x^2 + c*x^3 +d*x^4 + intercep

coef_ a*x^1	**coef_ b*x^2**	**coef_ c*x^3**	**coef_ d*x^4**	**intercept**
-5.49E+01	3.08E-01	-7.54E-04	6.84E-07	3567.28

- We checked the result manually by excel, then the result was correct.

3.5.3) Model conclusion:

As a result, we successfully built a polynomial regression to apply for correlation between vapor pressure and temperature of Butanol.
We can apply this method to analyze some other science field.
This method can help scientists to know the rules of the data better.

References

1. https://vi.wikipedia.org/wiki/Butan
2. Dortmund Data Bank Vapor Pressure of 1-Butanol http://www.ddbst.com/en/EED/PCP/VAP_C39.php
3. NIST Standard Reference Data https://webbook.nist.gov/cgi/cbook.cgi?ID=C71363&Mask=4&Type=ANTOINE&Plot=on
4. Polynomial Regression Pressure Dataset https://rpubs.com/anup_jana/polynomial
5. Difference Between Ideal Gas and Real Gas in Tabular Form - BYJU'S (byjus.com)
6. Tài liệu "Hướng dẫn làm thí nghiệm Quá trình & Thiết bị ".
7. Quá trình & Thiết bị trong công nghệ Hóa Học - Tập VII - Kỹ thuật Sấy vật liệu - Nguyễn Văn Lụa - ĐHBKTPHCM.
8. Sổ tay Quá trình & Thiết bị công nghệ Hóa Chất - Tập II - Chương VII - NXBKHKT - Hà Nội 1982.
9. Qúa trình và thiết bị công nghệ hóa học& thực phẩm.-Tập 3-Truyền khối. Võ Văn Bang, Vũ Bá Minh