APPENDIX

GAUSSIAN MIXTURE POLYMER VISUALIZATION

```
pd.set_option('display.float_format', lambda x: '%.5f' % x)
results = pd.DataFrame(np.array(list(zip(Feature, means.flatten(), mu.flatten(), var))), columns=['Polymer', 'Reference', 'Calculated Means', 'Calculated Variance'])
pd.set_option('display.float_format', lambda x: '%0.4f' % x)
table.append(results)

%mkdir graph$j

for i in range(len(mu)):
    # x value mean + stdv and mean - stdv
    x_axis=np.linspace(mu[i][0]-np.sqrt(var[i]),mu[i][0]+np.sqrt(var[i]), 10000)

     # for curve y value
     scaller = MinMaxScaler()
     y = scaller.fit_transform(norm.pdf(x_axis, mu[i][0], np.sqrt(var[i])).reshape(-1,1)).flatten()
     plt.plot(x_axis, y, label=i, lw=2)
     plt.title(Feature[i])

     # x_axis value on graph mean + stdv and mean - stdv
      plt.xticks(np.linspace(mu[i][0]-np.sqrt(var[i]), mu[i][0]+np.sqrt(var[i]), 5))
     plt.ticklabel_format(style='plain', useOffset=False)
     # plt.show()
     plt.savefig(f'graph%d/%02d - %s.jpg' % (j, i, Feature[i]))
     plt.clf()

# Save the graph
!zip -r graph.zip graph
```

DATA REGULARIZATION

```
def preprocessingTrain(importData):
    data_flat = importData.flatten()
    data_flat = data_flat[~np.isnan(data_flat)]
    data = data_flat.reshape(-1,1)
    return data
```

SPLIT X DATA

```
def data_x(p_x):
    x = []
    y = [1., 2., 3., 4., 5., 6., 7.]
    for dt in p_x:
        if dt[0] not in y:
            x.append(dt[0])
    return x
```

DATA AUGMENTATION

```
8. def crossover(A, B, ind):
9.     new = []
10.     for i in range(num):
11.         temp = []
12.
13.         len_a = A.shape[0]
14.         len_b = B.shape[0]
15.         maxi = min(len_a, len_b)-1
16.
17.         temp.append(ind)
18.         while len(temp) <= maxi:
19.             a1 = np.random.randint(0, len_a)
20.             b1 = np.random.randint(0, len_b)
21.
22.             temp.append(A[a1])
23.             temp.append(B[b1])
24.
25.         new.append(np.sort(temp))
26.     return new
```

PAPER NAME

TA-19.K1.0002.docx

WORD COUNT

5397 Words

CHARACTER COUNT

28266 Characters

PAGE COUNT

20 Pages

FILE SIZE

66.8KB

SUBMISSION DATE

Dec 14, 2022 2:07 PM GMT+7

REPORT DATE

Dec 14, 2022 2:08 PM GMT+7

5% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

- 3% Internet database
- Crossref database
- 4% Submitted Works database
- 1% Publications database
- Crossref Posted Content database