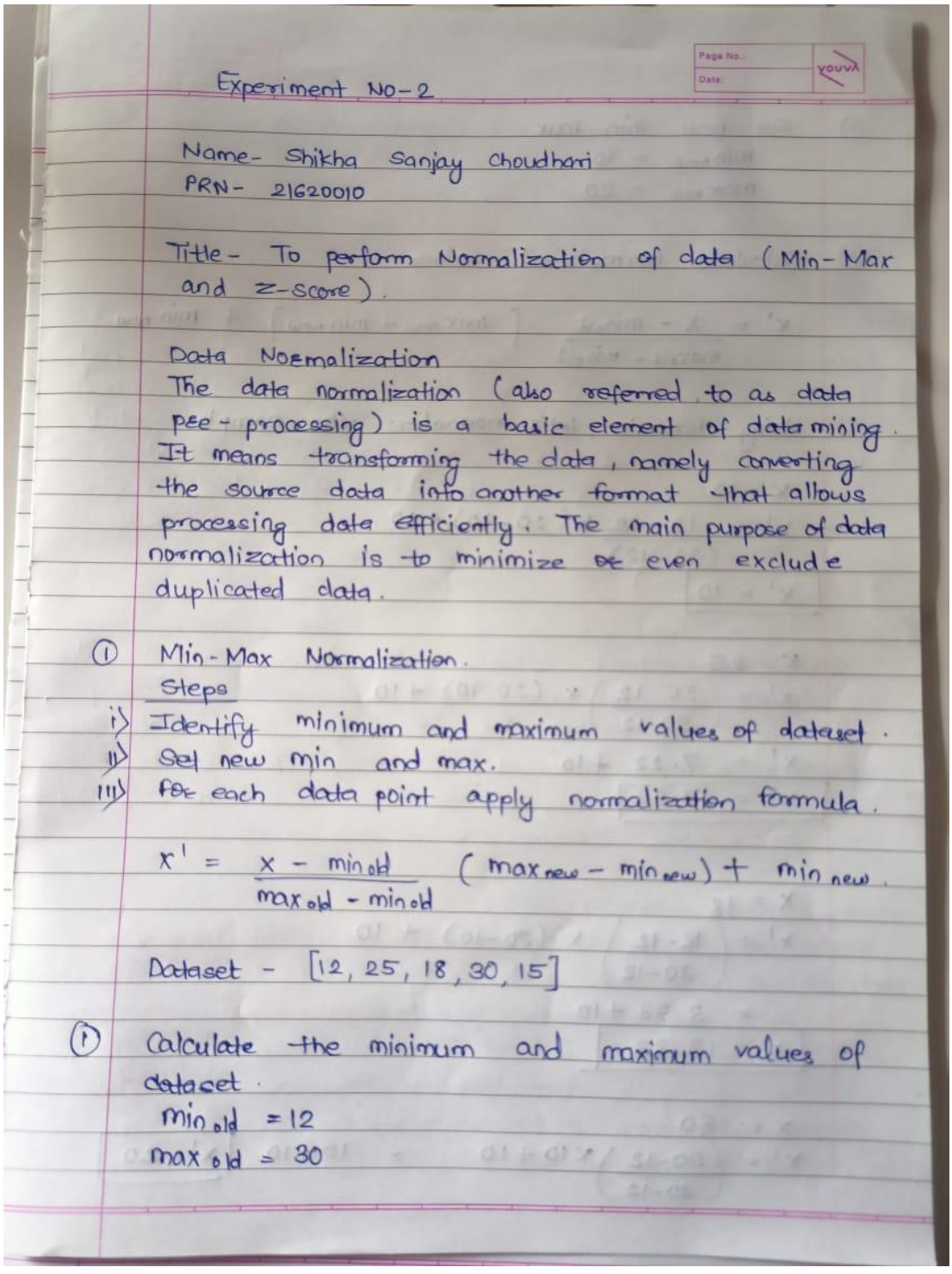
Experiment 2

Name - Shikha Sanjay Choudhari
PRN- 21620010

Title- To perform normalization of data (Min-Max and Z-score)



Experiment No-2

Name- Shikha Sanjay Choudhari
PRN- 21620010

Title- To perform Normalization of data (Min-Max and z-score).

Data Normalization
The data normalization (also referred to as data pre-processing) is a basic element of data mining. It means transforming the data, namely converting the source data into another format that allows processing data efficiently. The main purpose of data normalization is to minimize & even exclude duplicated data.

① Min-Max Normalization.

Steps
i) Identify minimum and maximum values of dataset.
ii) Set new min and max.
iii) For each data point apply normalization formula.

$$x' = \frac{x - min_{old}}{max_{old} - min_{old}} (max_{new} - min_{new}) + min_{new}.$$

Dataset - [12, 25, 18, 30, 15]

① Calculate the minimum and maximum values of dataset.

$min_{old} = 12$
$max_{old} = 30$

② Set new min max.

$min_{new} = 10$

$max_{new} = 20$

③ Calculate normalized values.

$$x' = \frac{x - min_{old}}{max_{old} - min_{old}} [max_{new} - min_{new}] + min_{new}.$$

Using formula, let's normalize the sample.

$x = 12$

$$x' = \frac{12-12}{(30-12)} \times (20-10) + 10$$

$$\boxed{x' = 10}$$

$x = 25$

$$x' = \left(\frac{25-12}{30-12}\right) \times (20-10) + 10$$

$$x' = 7.22 + 10$$

$$\boxed{x' = 17.22}$$

$x = 18$

$$x' = \left(\frac{18-12}{30-12}\right) \times (20-10) + 10$$

$$= 3.33 + 10$$

$$\boxed{x' = 13.33}$$

$x = 30$

$$x' = \left(\frac{30-12}{30-12}\right) \times 10 + 10 = 10 + 10 \quad \boxed{x' = 20}$$

$x = 15$

$$x' = \left(\frac{15-12}{30-12}\right) \times (10-20) + 10$$

$$x' = 1.66 + 10$$

$$\boxed{x' = 11.66}$$

So, the normalized data with min-max normalization is -

$$[10, 17.22, 13.33, 20, 11.66]$$

(2) z-score normalization.

Steps -

1) Calculate the mean and standard deviation of the sample data.
2) For each data point, calculate the z-score using formula -

$$x' = \frac{x - \text{mean}}{\text{standard deviation}}.$$

Let's perform z-score normalization

Sample data - $[12, 25, 18, 30, 15]$

(1) Calculate mean and S.D.

① Mean - $\frac{12+25+18+30+15}{5}$

$$\boxed{\bar{x} = 20}$$

Standard Deviation

$$= \sqrt{\frac{(12-20)^2 + (25-20)^2 + (18-20)^2 + (30-20)^2 + (15-20)^2}{5}}$$

$$= 6.16$$

(2) Calculate z-score.

$X = 12 \quad x' = \frac{12-20}{6.16} = -1.30$

$X = 25 \quad x' = \frac{25-20}{6.16} = 0.81$

$X = 18 \quad x' = \frac{18-20}{6.16} = -0.32$

$X = 30 \quad x' = \frac{30-20}{6.16} = 1.62$

$X = 15 \quad x' = \frac{15-20}{6.16} = -0.81$

z-scored normalized data would be -

$[-1.30, 0.81, -0.32, 1.62, -0.81]$

- **Code**

```cpp
#include <bits/stdc++.h>
#include <fstream>
#include <vector>
#include <cmath>
#include <string>

using namespace std;

// Function to perform min-max normalization using the specified formula
vector<double> minMaxNormalization(const vector<double>& data, double min_new,
double max_new) {
    double min_old = *min_element(data.begin(), data.end());
    double max_old = *max_element(data.begin(), data.end());

    vector<double> normalized_data;
    for (double val : data) {
        double normalized_val = (val - min_old) / (max_old - min_old) * (max_new
- min_new) + min_new;
        normalized_data.push_back(normalized_val);
    }

    return normalized_data;
}

// Function to perform z-score normalization
vector<double> zScoreNormalization(const vector<double>& data) {
    double mean = 0;
    for (double val : data) {
        mean += val;
    }
    mean /= data.size();

    double variance = 0;
    for (double val : data) {
        variance += (val - mean) * (val - mean);
    }
    variance /= data.size();

    double std_deviation = sqrt(variance);

    vector<double> normalized_data;
    for (double val : data) {
        double z_score = (val - mean) / std_deviation;
        normalized_data.push_back(z_score);
```

```
    }

    return normalized_data;
}

int main() {
    ifstream input_file("input.txt");
    ofstream output_file("output.txt");

    vector<double> data;
    double value;

    while (input_file >> value) {
        data.push_back(value);
    }

    double min_new, max_new;
    cout << "Enter the new range for min-max normalization (min new max): ";
    cin >> min_new >> max_new;

    vector<double> min_max_normalized = minMaxNormalization(data, min_new,
max_new);
    vector<double> z_score_normalized = zScoreNormalization(data);

    output_file << "Min-Max Normalized Data:\n";
    for (double val : min_max_normalized) {
        output_file << val << "\n";
    }

    output_file << "\nZ-Score Normalized Data:\n";
    for (double val : z_score_normalized) {
        output_file << val << "\n";
    }

    input_file.close();
    output_file.close();

    return 0;
}
```

- **Input data**

input.txt

```
1  //Input data
2
3  12
4  25
5  18
6  30
7  15
8
```

- **Output**

output.txt

```
1  Min-Max Normalized Data:
2  10
3  17.2222
4  13.3333
5  20
6  11.6667
7
8  Z-Score Normalized Data:
9  -1.21157
10 0.757228
11 -0.302891
12 1.51446
13 -0.757228
14
```