FPE Library Documentation

1. Library Overview

The **Feature Probability-based Estimation (FPE)** library helps identify the most important features in a dataset by calculating feature-wise probabilities. It is particularly useful in feature selection tasks to improve machine learning model performance by removing less significant features.

Dataset Conditions

1. **Data Consistency**: Each feature must have data that is either entirely made up of strings or entirely numeric values. Mixing different types is not allowed.
2. **No Missing Data**: No empty cells (e.g., NaN or blanks) should be in the feature data. Every cell must contain a value.
3. **Uniform Data Type**: All values within a single feature must follow the same data type. For example, if the feature is numeric, all values should be numbers.
4. **Target (Label) Column**: The last column contains the target or Label of the dataset. It works on labeled datasets.

The following Python libraries are required for the FPE execution, and these can be installed using the commands below:

1. **Pandas** - Install using: pip install pandas
2. **NumPy** - Install using: pip install numpy (***Optional***)
3. **Numbers -** This is a built-in Python module; if required**,** (*pip install numbers*).
4. **warnings** - This is also a built-in Python module, if required**,** (*pip install warnings*).

---

2. Installation

Install the FPE library using pip:

```
pip install fpe-lib==0.1.2
```

---

3. Usage

3.1 Importing the Library

After installation, import the library:

```
from fpe.fpe import fpefs
```

3.2 Input Dataset

- **Features (X)**: All columns except the last one are considered features.
- **Target (y)**: The last column is treated as the target variable.

3.3 Example Usage

```
import pandas as pd
from fpe.fpe import fpefs

# Sample dataset
data = pd.DataFrame({
    'Feature1': [1, 2, 3, 4, 5],
    'Feature2': ['A', 'B', 'A', 'B', 'C'],
    'Target': [1, 0, 1, 0, 1]
})

# Apply FPEFS
result = fpefs(data)

# View results
print(result)
```

Output:

```
  Feature  Probability
0  Feature1         0.75
1  Feature2         0.75
```

References/Resources:

[1] Prasad, Mahendra, Sachin Tripathi, and Keshav Dahal. "An efficient feature selection based Bayesian and Rough set approach for intrusion detection." *Applied Soft Computing* 87 (2020): 105980.

[2] Prasad, Mahendra, Sachin Tripathi, and Keshav Dahal. "A probability estimation-based feature reduction and Bayesian rough set approach for intrusion detection in mobile ad-hoc network." *Applied Intelligence* 53.6 (2023): 7169-7185.

[3] Prasad, Mahendra, Sachin Tripathi, and Keshav Dahal. "A Feature Probability Estimation-based Feature Selection Approach for Intrusion Detection." *RAIT-2025*.

[4] Data Repository: https://github.com/mahendrapd/FPE