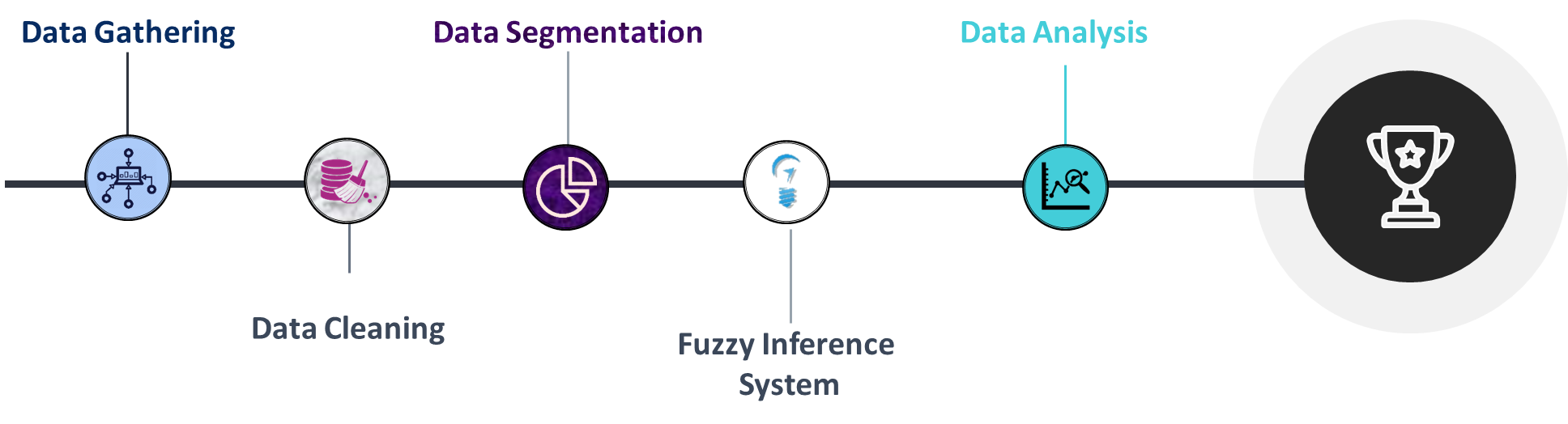
**Design**

****

**Data Gathering**

* This step involves the extraction of data from the Source.
* Then, storing the data collected into a pre-defined format.
* Data is stored in a structured format

**Data Cleaning**

* It is also called the Preprocessing stage.
* This step involves removing redundant or error-prone data from the dataset.
* This process consists of handling NULL, NAN and any other odd values.
* Typecasting ( to specify the data in user-defined type and to maintain consistency ) and Data normalization ( If the comparison is needed to be done between two different features ) are also the part of this phase.
* This is one of the crucial phases and directly influence the outcome of the process. Without pre-processing, data can result in unexpected results while results obtained from pre-processed data are more consistent and reliable.

**Data Segmentation**

* This step involves Segmentation of data on the basis of similarity of features.
* This process will decompose the whole dataset into segments in which each data follows a general trend in one or more features.
* In this phase, we are implementing four segmentation algorithm (Top-Down, Bottom-Up, Sliding Window, Dual Sliding Window).
* Each Algorithm works on a different set of hyper parameters e.g. sliding window algorithm needs window size and threshold while top down algorithm requires the Threshold only.

**Fuzzy Inference System**

* There is no crisp decision boundary to differentiate between modes of Transportation, thus this project uses the concept of Fuzzy Inference system.
* Fuzzy Inference System uses a Rule base that specifies a set of If- Else conditions to determine the final Crisp Output and an inference engine that evaluates the values on the basis of different membership functions.

**Data Analysis**

This step involves multiple subtasks:

* Identification of optimal threshold for segmentation.
* Identification factor (or factors) of classification of Mode of transportation.
* Comparison of different segmentation algorithms on different factors like time complexity, space complexity, accuracy etc.

**Modules Used**

**Python Modules:-**

1. **Pandas:**

In PYTHON programming, Pandas is a software library written for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and time series. It is free 17 software released under the three-clause BSD license.

1. **Numpy**

**Numpy**is a general-purpose array-processing package. It provides a high-performance multidimensional array object and tools for working with these arrays. It is the fundamental package for scientific computing with Python.

1. **Matplotlib**

Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hardcopy formats and interactive environments across platforms

1. **Folium**

Folium builds on the data wrangling strengths of the Python ecosystem and the mapping strengths of the Leaflet.js library. Manipulate your data in Python, then visualize it in on a Leaflet map via Folium.

1. **Skfuzzy**

This package implements many useful tools for projects involving fuzzy logic, also known as grey logic. It is pure Python and depends only on NumPy, SciPy, and NetworkX. There are many basic fuzzy logic functions in the base namespace, as well as a Pythonic, object-oriented system for fuzzy control systems in the skfuzzy.control submodule.

1. **Preprocessing**

* Time\_Extractor.py
* Distribution\_calculator.py

1. **Data Cleaning**

* Preproccessing.py

1. **Data Segmentation**

* **Dual Sliding Window**

Two fixed sized windows are created adjacent to each other. The distribution of data of both windows are then compared if the difference exceeds the limit specified by the user, then the anchor is placed between the windows.

* **Sliding Window**

The anchor is placed when the difference created by including one more data point is more than the predefined threshold.

* **Top-Down**

The anchor is placed when the difference between the variance of subsequent segments is more than a user-defined threshold. The segments are computed in the Divide and Conquer manner.

* **Bottom-Up**

The anchor is removed when the difference between means of consecutive segments is less than the user-specified threshold.

1. **Fuzzy Inference System**

* FuzzySystem.py (under Development)

1. **Data Analysis**

* Performance.py (under Development)