**Problem statement:** Analyzing changes in signal data occur due to multipath effects in both indoor and outdoor environments.

The main objectives are as follows:

1. To provide a continuous processing system (CPS) by using Spark SQL engine for structural streaming data.
2. To identify new changes in time series that occurs for streaming data.

**Reported outcomes:** Prerequisites for Apache spark streaming in Jupyter python notebook are implemented.

**Specific Aims:**

1. Downloaded apache spark Hadoop for streaming analytics in same folder where python files are stored.
2. Downloaded and installed anaconda for implementation.
3. Jupyter notebook has been launched and relevant packages are installed for data exploration analysis.
4. Gather all related papers and categorize them.

**Key Accomplishments:**

1. Latest version of Java and spark Hadoop are downloaded and installed.
2. Required packages for apache spark streaming are imported in anaconda power shell prompt.
3. Loaded the BLE Dataset by using Pandas DataFrame library in python.
4. Implemented the plots for BLE Dataset by using matplotlib library in python.

**Red Flags:**

1. We are not handling streaming data.
2. We are not data-driven.
3. Error Occurred: Need to convert the location column to float from string in BLE dataset.
4. Need some more time for preprocessing the BLE dataset, which is not yet completed this week.
5. Need to identify how apache spark streaming is integrated with Kafka (input source) for generating beacon data.

**Future Work:**

1. To implement apache spark streaming with Kafka integration, so that it can build an application for generating beacon signal data.

**Timeline (tentative timeline for the upcoming week)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Future Specific aims** | **10/02** | **10/03** | **10/04** | **10/07** | **10/08** | **10/09** |
| Pre-processing of BLE dataset |  |  |  |  |  |  |
| Identifying the approach for Apache Spark Streaming with Kafka Integration |  |  |  |  |  |  |
| Implementation of change detection based on non-parametric approaches |  |  |  |  |  |  |

**References:**

[1] S. Aminikhanghahi and D. J. Cook, “A survey of methods for time series change point detection,” *Knowledge and Information Systems*, vol. 51, no. 2, pp. 339–367, Aug. 2016.

[2] A. Bifet, G. Holmes, and B. Pfahringer, “MOA-TweetReader: Real-Time Analysis in Twitter Streaming Data,” *Discovery Science Lecture Notes in Computer Science*, pp. 46–60, 2011.

[3] “Analysis of real-time data with spark streaming,” *Journal of Advances in Technology and Engineering Research*, vol. 3, no. 4, 2017.

**Appendix A**

**Apache Spark Streaming**

**PREREQUISITES:**

1. Setting path for latest Java version (1.8.0) and spark Hadoop (2.4.4) are done in same folder, where Jupyter python files are saved.
2. A list of packages are imported for apache spark streaming:
   1. pyspark: Python API written in python to support Apache Spark. It requires python to be available on the system path.
   2. findspark: It helps to find the location of apache spark Hadoop.
   3. SparkContext: It is the entry point to any spark functionality.
   4. StreamingContext: It represents the connection to a spark cluster and can be used to DStream various input sources such as Kafka, Flume and Kinesis.
   5. SQLContext: It is the entry point into all SQL functionality in spark.

