**BIG – DATA Project**

**ON**

**Health Monitoring System**

**By**

**Abhi Raj**

**12212645**

**9W085**

****

**Health Monitoring System Project Report**

**Introduction**

The Health Monitoring System project aims to analyze health data of 10,000 patients collected from a diagnostic center. The project uses big data technologies such as PySpark, visualization libraries like Seaborn and Plotly, and interactive dashboards built with Dash.

**Objective**

The primary objective of this project is to process large-scale health data using Apache Spark and generate statistical insights into patients' health parameters like Blood Pressure (BP), Sugar Level, Cholesterol, and Haemoglobin.

**Tools and Technologies Used**

* **PySpark:** For distributed data processing
* **Pandas:** For data manipulation
* **Seaborn & Matplotlib:** For static data visualizations
* **Plotly:** For interactive visualizations
* **Dash:** For interactive dashboard creation
* **Google Colab:** For code execution and cloud-based development

**Data Description**

The dataset contains the following columns:

* Patient\_ID
* BP (Blood Pressure)
* Sugar\_Level
* Cholesterol
* Haemoglobin

**Sample Data**

| **Patient\_ID** | **BP** | **Sugar\_Level** | **Cholesterol** | **Haemoglobin** |
| --- | --- | --- | --- | --- |
| 1001 | 120 | 90 | 180 | 14 |
| 1002 | 130 | 110 | 200 | 13.5 |
| 1003 | 140 | 100 | 190 | 14.2 |

**Data Processing**

1. **Data Loading:** The CSV file containing patient health data is uploaded into Google Colab.
2. **Spark DataFrame Conversion:** The dataset is converted into Spark DataFrame for efficient data processing.
3. **Data Overview:** The schema and initial records are displayed to understand the dataset structure.
4. **Data Aggregation:** Average values for BP, Sugar Level, Cholesterol, and Haemoglobin are calculated using PySpark.

**Data Analysis**

The average health parameters were calculated as follows:

| **Parameter** | **Average Value** |
| --- | --- |
| BP | 128 |
| Sugar\_Level | 100 |
| Cholesterol | 190 |
| Haemoglobin | 13.9 |

**Visualization**

The average health parameters were visualized using Seaborn and Plotly. A bar plot displays the comparative average values of each health parameter.

**Dashboard**

An interactive dashboard was created using Dash, which allows users to view the statistical summary graphically. The dashboard includes:

* Title Header
* Interactive Bar Chart

**Conclusion**

The Health Monitoring System successfully processed large-scale patient health data, generated statistical insights, and provided an interactive dashboard for data visualization. This project demonstrates the application of big data technologies in healthcare analytics.

**Future Enhancements**

* Adding more patient parameters
* Predictive analysis for patient risk categorization
* Real-time data streaming and analysis

**References**

* Apache Spark Documentation
* Plotly Official Documentation
* Dash Official Documentation
* Google Colab Documentation