**MOHAMED SATHAK A.J.COLLEGE OF ENGINEERING**

**COLLEGE CODE-3118**

**PROJECT TITLE: BIG DATA ANALYSIS WITH IBM CLOUD DATABASES**

**TEAM MEMBERS:**

**GULABI BASHARATH.A-311821104011**

**SWETHA.S-311821104051**

**UMA MAHESWARI.E.D-311821104059**

**DEVI.G-311821104007**

**SYEDA ZUHA TASNEEM-311821104052**

**Phase 1: Problem Definition and Design Thinking**

**Problem Definition**

* Clearly define the scope of your big data analysis project. Determine the specific datasets you want to analyse on social media trends.
* Establish the goals and objectives of the analysis. What insights are you looking to gain, and how will they benefit your organization or project?
* Define the timeline and budget for the project..

**Data Selection**

* Define Research Focus: Clearly outline your research objectives and the specific social trend you want to study.
* Select Reliable Data Sources: Choose trustworthy and relevant data sources that provide insights into the chosen social trend.
* Analyze & Convey Insights: Apply appropriate analysis techniques to the selected data. Effectively communicate your findings to convey the trends and patterns observed in the social context.

**Database Setup**

* Sign up for an IBM Cloud account if you don't already have one.
* Create an IBM Cloud Database instance suitable for storing and managing large datasets. Choose the appropriate database technology (e.g., Db2, PostgreSQL) based on your project's requirements.
* Import the selected datasets into the database. Ensure proper indexing and organization of data for efficient querying.

**Data Exploration**

* **Data Visualization**: Create visual representations of your data, like charts and graphs, to identify patterns and trends.
* **Statistical Analysis**: Use statistical techniques to uncover relationships, correlations, and significant findings within your data.
* **Interpret and Communicate**: Carefully interpret your findings and effectively communicate insights through reports and visualizations to better understand the social trend.

**Analysis Techniques**

* Statistical analysis: Descriptive statistics, hypothesis testing, correlation analysis, etc.
* Machine learning: Classification, regression, clustering, or deep learning, depending on the problem.
* Time series analysis if you're working with temporal data.
* Train and validate your models if you are using machine learning techniques.
* Preprocess the data as needed for the selected analysis techniques (e.g., feature engineering, data transformation).
* Implement the chosen analysis methods and algorithms, ensuring proper validation and testing.

**Visualization**

* Design visualizations that effectively communicate the results of your analysis. Consider using tools like Python's Matplotlib, Seaborn, or Tableau.
* Create informative charts, graphs, and dashboards to present your findings in a visually appealing and understandable way.

**Business Insights**

* Interpret the analysis results to derive valuable business insights. Address the original objectives and goals defined in the problem definition phase.
* Identify actionable recommendations based on the insights obtained. These recommendations should be tailored to the specific needs of your organization or project.

**Documentation and Reporting**

* Document your entire analysis process, including data sources, methods, code, and results.
* Create a comprehensive report or presentation that summarizes your findings and recommendations for stakeholders.

**Iterate and Refine**

* Review and refine your analysis based on feedback and new insights.
* Consider automating parts of the analysis process for future use, especially if this is an ongoing project.

**Deployment and Monitoring**

* If applicable, deploy any predictive models or automated analysis pipelines into production.
* Set up monitoring and alerts to ensure that the analysis continues to provide value over time.

**Conclusion**

* Remember that big data analysis is an iterative process, and continuous improvement is key to deriving valuable insights. Keep refining your analysis techniques and data strategies as you learn from each project iteration