Big data analysis with IBM cloud databases

PHASE-2:-

ADVANCED ML FOR BIG DATA:-

Incorporating advanced machine learning algorithms for predictive analysis or anomaly detection in big data can be highly beneficial. Some powerful algorithms to consider include:

1. **Random Forest:** Effective for both classification and regression tasks, providing robust predictions and feature importance analysis.

2. **Gradient Boosting:** Algorithms like XGBoost, LightGBM, and CatBoost can boost predictive accuracy by combining weak models.

3. **Neural Networks:** Deep learning models, such as Convolutional Neural Networks (CNNs) for image data and Recurrent Neural Networks (RNNs) for sequential data, can handle complex patterns.

4. **Support Vector Machines (SVM):** Useful for binary classification and applicable to various domains.

5. **Clustering Algorithms:** Techniques like K-Means or DBSCAN can help identify anomalies through unsupervised learning.

6. **Time Series Analysis:** Algorithms like ARIMA, Prophet, or LSTM can model and predict time-dependent data.

7. **Isolation Forest:** Ideal for outlier detection in high-dimensional data.

8. **Principal Component Analysis (PCA):** Helpful for dimensionality reduction and feature engineering.

9. **AutoML:** Tools like Google AutoML, H2O.ai, or Auto-sklearn automate model selection and hyperparameter tuning.

When implementing these algorithms for big data, consider distributed computing frameworks like Apache Spark or cloud-based platforms that can handle the volume and velocity of data. Additionally, feature engineering, data preprocessing, and model evaluation are crucial aspects of the process.

TEAM MEMBERS:-

NAGAPPAN

AKASH

SIVAKARTHICK

SANDEEP

BARATH KUMAR