

ABSTRAK

Curah hujan merupakan salah satu parameter klimatologi yang memiliki peran penting dalam mendukung sektor pertanian, pengelolaan sumber daya air, dan mitigasi bencana hidrometeorologi di Indonesia. Seiring dengan meningkatnya volume data meteorologi yang dihasilkan secara kontinu, diperlukan teknologi yang mampu mengolah data dalam jumlah besar secara efisien. Penelitian ini bertujuan untuk menganalisis pola curah hujan di Indonesia menggunakan pendekatan Big Data dengan memanfaatkan Apache Spark sebagai framework pengolahan data. Data yang digunakan berupa data curah hujan harian dari 34 provinsi di Indonesia yang diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) dalam format CSV. Metode penelitian yang digunakan adalah kuantitatif deskriptif dengan tahapan meliputi data ingestion, data cleaning, transformasi data, dan analisis statistik deskriptif menggunakan Apache Spark berbasis PySpark. Analisis dilakukan untuk memperoleh informasi mengenai jumlah data valid, nilai rata-rata, nilai maksimum, dan nilai minimum curah hujan pada setiap provinsi. Hasil penelitian menunjukkan bahwa Apache Spark mampu mengolah data curah hujan secara terstruktur, efisien, dan otomatis. Berdasarkan hasil analisis, Provinsi Sumatera Barat memiliki rata-rata curah hujan tertinggi sebesar 16,01 mm/hari, sedangkan Provinsi Sulawesi Tengah memiliki rata-rata curah hujan terendah sebesar 0,30 mm/hari. Selain itu, curah hujan maksimum tertinggi tercatat di Provinsi Bengkulu sebesar 186,4 mm. Penggunaan Spark DataFrame serta fungsi agregasi seperti *count()*, *avg()*, *max()*, dan *min()* terbukti mampu mendukung proses pengolahan dan analisis data secara efektif. Hasil penelitian menunjukkan bahwa Apache Spark dapat digunakan sebagai solusi yang efektif dalam pengolahan data klimatologi berskala besar dan mampu menghasilkan informasi yang bermanfaat untuk mendukung pengambilan keputusan dalam bidang pertanian, pengelolaan sumber daya air, serta mitigasi bencana di Indonesia.

ABSTRACT

Rainfall is one of the most important climatological parameters that plays a significant role in supporting agriculture, water resource management, and hydrometeorological disaster mitigation in Indonesia. As the volume of meteorological data continues to increase, technologies capable of processing large-scale data efficiently are required. This study aims to analyze rainfall patterns in Indonesia using a Big Data approach by utilizing Apache Spark as the primary data processing framework. The data used in this research consist of daily rainfall data from 34 provinces in Indonesia obtained from the Meteorology, Climatology, and Geophysics Agency (BMKG) in CSV format. This research employs a descriptive quantitative method consisting of data ingestion, data cleaning, data transformation, and descriptive statistical analysis using Apache Spark with PySpark. The analysis was conducted to obtain information regarding the number of valid records, average rainfall, maximum rainfall, and minimum rainfall values for each province. The results indicate that Apache Spark is capable of processing rainfall data in a structured, efficient, and automated manner. Based on the analysis, West Sumatra Province recorded the highest average rainfall at 16.01 mm/day, while Central Sulawesi Province recorded the lowest average rainfall at 0.30 mm/day. In addition, the highest maximum rainfall was observed in Bengkulu Province at 186.4 mm. The use of Spark DataFrame and aggregation functions such as *count()*, *avg()*, *max()*, and *min()* proved effective in supporting data processing and analysis activities. The findings demonstrate that Apache Spark can serve as an effective solution for large-scale climatological data processing and provide valuable information to support decision-making in agriculture, water resource management, and disaster mitigation in Indonesia.