Abstract

In the age of data-driven decision-making, our project dives into the world of real-time data analysis, focusing on music streaming. By employing simulated data, we replicate the complexities of online music consumption, including user interactions, genre preferences, and playlist dynamics, enabling us to develop and evaluate advanced analytics solutions tailored to the music industry. Through the integration of Apache Kafka, Apache Spark, MongoDB, and cutting-edge machine learning techniques, we construct an end-to-end data processing pipeline capable of handling large-scale streaming data. This project explores data generation, storage, analysis, and visualization, providing valuable insights into listener behaviour, music consumption patterns, and the effectiveness of various promotional strategies.

By examining the scalability and adaptability of our analytical framework, we offer insights into the future of real-time data analysis in the dynamic and rapidly evolving landscape of music streaming services. Our findings not only contribute to advancing the understanding of user behaviour in the digital music domain but also pave the way for innovative solutions to address challenges and opportunities in the industry.